QUALITY CONTROL CONVENTION PAPERS 1954

J. S. Herz

EIGHTH ANNUAL CONVENTION

AMERICAN SOCIETY
FOR
QUALITY CONTROL

QUALITY CONTROL CONVENTION PAPERS 1954

EIGHTH ANNUAL CONVENTION

AMERICAN SOCIETY FOR QUALITY CONTROL

JUNE 9, 10, 11, 1954

ST. LOUIS, MISSOURI

PRICE

\$3.00 at the Convention

\$3.50 if Ordered by Mail

Copyright 1954, By
AMERICAN SOCIETY FOR QUALITY CONTROL, INC.
Room 5736
70 East 45th Street
New York 17, New York

Printed in the United States of America By
THE ENGLE PRESS
Philadelphia, Pa.

FOREWORD

The excellent cooperation of the authors in meeting an early deadline and in presenting their manuscripts in suitable form for direct reproduction has made this publication possible.

Of the sixty-seven (67) sessions scheduled for the Convention sixty-two (62) involved papers potentially available for these Transactions. Of these, sixty-one (61)* are presented here. Due to illness Dr. Irving W. Burr was unable to prepare his paper in time for inclusion here. Copies will be distributed at the Convention and anyone else interested in obtaining a copy should consult page 625.

Lt. Col. Ralph M. Lockhardt (I-B-1) was unable to present his paper here since the program his talk covers has been undergoing revision and could not be properly described prior to Convention time.

Session III-B-2 is a panel presentation which is informal and from which no material is available for publication.

One speaker had to withdraw at a late date and time was not available for a replacement before press time (IV-E-1).

Prof. J. A. Henry has made plans to prepare a separate publication of the sessions on Inplant Training (IV-A). This publication is to include written discussions of the papers presented and the transcribed discussions at the Convention. Those interested in obtaining copies should write to Prof. J. A. Henry, Department of Mechanical Engineering, University of Illinois, Urbana, Illinois.

While these Transactions are copyrighted, The American Society for Quality Control assumes no responsibility for any of the authors' statements. Responsibility for the content of each paper resides with its author.

The Editor is greatly indebted to Mr. John G. Rutherford, Program Chairman of The General Convention Committee, and Mr. Lester S. Kauffman, Transactions Editor of The St. Louis Convention Committee for their assistance in gathering the material for The Transactions and for sale and distribution at the Convention.

Edward M. Schrock

Edward M. Schrock
Transactions Editor
General Convention Committee

*Actually this has turned out to be sixty (60) since one paper expected at the last moment before going to press has apparently been lost in transit.

Ed.

NATIONAL OFFICERS

ALFRED L. DAVIS, President
Rochester Institute of Technology

PAUL A. ROBERT, Vice President
International Business Machines Corporation

RAYMOND S. SADDORIS, Vice President
A. O. Smith Corporation

JULIAN H. TOULOUSE, Vice President
Owens-Illinois Glass Company

DORIAN SHAININ, Executive Secretary
Rath and Strong, Incorporated

LEON BASS, Treasurer
General Electric Company

MASON E. WESCOTT, Chairman, Editorial Board
Rutgers University

SIMON COLLIER, Jr. Past President
Johns-Manville Corporation

ST. LOUIS SECTION OFFICERS

DONALD F. SAGER, Chairman
Anheuser-Busch, Incorporated

RUDOLPH FREEDMAN, Vice Chairman
Alco Valve Company

THAINE O. JOHNSON, Secretary
Falstaff Brewing Corporation

GEORGE D. DREW, Treasurer
Gaylord Container Corporation

CONVENTION COMMITTEES
8th ANNUAL A.S.Q.C. CONVENTION

General Chairman
PAUL K. LEATHERMAN, Lambert Pharmacal Company

Vice Chairman
C. E. FISHER, Bell Telephone Laboratories

Secretary-Treasurer
ARNOLD G. BRAUN, Carter Carburetor Corporation

Arrangements
KORT K. PFABE, Carter Carburetor Corporation
E. A. BOYER, Carter Carburetor Corporation

Book Exhibits
JOSEPH MOVSHIN, Ramsey Corporation

Exhibits
FRANCIS B. FISHER, Century Electric Company
ALAN W. PETERSON, Century Electric Company

Transactions
LESTER S. KAUFFMAN, Magic Chef, Inc.

Program
LLOYD A. KNOWLER, State University of Iowa

Publicity and Promotion
RUDOLPH FREEDMAN, Alco Valve Company
MILTON WEISS, Alco Valve Company

Reception
DONALD F. SAGER, Anheuser-Busch, Inc.

Registration
ULYSSES C. GRAMSCH, Falstaff Brewing Corp.
THAINE O. JOHNSON, Falstaff Brewing Corp.

Special Events
HENRY O. SCHOTT, Monsanto Chemical Company

GENERAL CONVENTION COMMITTEE - A.S.Q.C.

WADE R. WEAVER, Chairman, Republic Steel Corp., Cleveland Ohio
FRANK CAPLAN, JR., General Electric Co., Syracuse, N. Y.
CHARLES R. KENDEL, Lamson & Sessions Co., Cleveland, Ohio
PAUL K. LEATHERMAN, Lambert Pharmacal Co., St. Louis, Mo.
WILLIAM J. MASSER, General Electric Co., Philadelphia, Pa.
ELLIS R. OTT, Rutgers University, New Brunswick, N. J.
JOHN G. RUTHERFORD, Lear, Incorporated, Grand Rapids, Michigan
ERVIN E. SCHIESEL, The Mattatuck Mfg. Co., Waterbury, Conn.
EDWARD M. SCHROCK, Koppers Company, Inc., Baltimore, Maryland

TABLE OF CONTENTS

F. CHEMICAL

G. COMPLEX ASSEMBLIES

H. ELECTRONICS

L. NEW STATISTICAL DEVELOPMENTS

M. OPERATIONS RESEARCH

N. PAPER AND PRINTING

O. PRECISION MANUFACTURING

LISTING BY AUTHORS

NOTES

NOTES

QUALITY CONTROL IN THE MANUFACTURE OF ZINC
TECHNIQUE OF APPROXIMATING A 3-VARIABLE PROBLEM SOLUTION

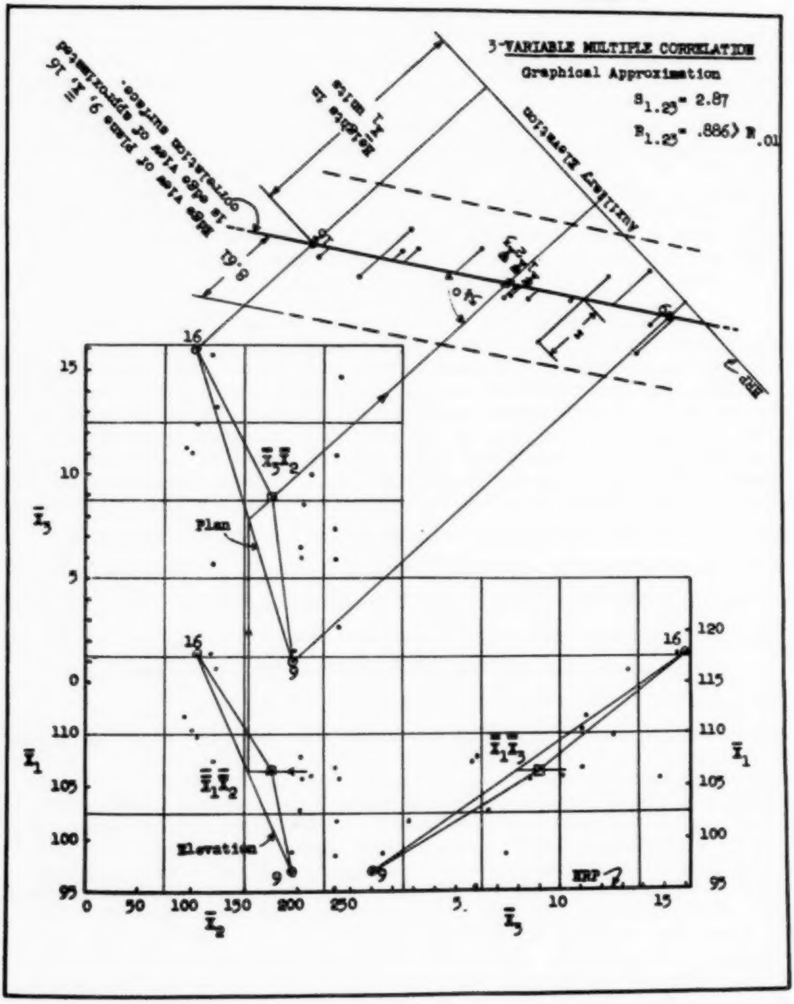
Harold L. Springer
St. Joseph Lead Company
Zinc Smelting Division

The Quality Control Engineers who have had occasion to carry through linear multivariate regression solutions know from experience that often many variables are treated in order that the significant independent variates may be isolated. Also, quite frequently the number of independent variables interacting significantly at a given time turns out to be two, three, or possibly four. Because time is at a premium and because complete calculations are lengthy, many multivariate analyses are only partially solved or perhaps remain unanalyzed. The purpose of this paper is to suggest a means of relief from this situation and at the same time provide a visual development of the problem which should help clarify the concept of multivariate regression. The relation of the concept of multivariate regression to analysis of control chart data is also examined.

In most statistical textbooks it will be seen that a 3-variate multiple correlation solution involves a regression PLANE just as a simple regression problem involves a LINE. If we apply our basic engineering training in descriptive geometry we find we can rapidly approximate the solution of a 3-variate correlation problem. A plot of three variables arranged in the form of the ordinary orthographic projection system where the top, front, and end views are normal to each other, provides three simple regression line graphs. It is not necessary to plot the end view because we actually only need to use the plan view, which is a plot of the independent variates $\overline{X}_3$ on $\overline{X}_2$; and the elevation, which is a plot of the dependent variate $\overline{X}_1$ on $\overline{X}_2$. Positions of $\overline{X}_2$ and $\overline{X}_3$ may be interchanged without disturbing the solution. The bar X's are used for purposes which will be obvious later.

At the outset we know (by least squares) that the grand mean $\overline{\overline{X}}_1\overline{\overline{X}}_2\overline{\overline{X}}_3$ will lie on the regression plane. For the sake of simplicity, samples 9 and 16 are used with the grand mean in order to define a plane that approximates the least squares solution. This is shown on Graph I. However, in most cases it will be more expedient to reserve the grand mean for a check and use instead three or four averages to determine the plane. Each of these averages should be made up of several sample values from the same general location with respect to a given plot, and, the averages should be chosen from the plot (plan or elevation) that appears to show the lesser correlation in order that the spread between the averages will be as wide as the data will permit. This will tend to speed the plotting by reducing the variation in the data and, consequently, the geometry. It should be noted that a warped surface will result when one or more of the variables is non-linear and the proper transformation should be made before the plotting is done. A plot of three or more points along with the grand mean will produce a polygon instead of a triangle and will, in general, assure greater accuracy. As for the geometry, when we pass a horizontal plane through $\overline{\overline{X}}_1\overline{\overline{X}}_2$ and cut the line 9, 16 we establish the direction in which to draw the

GRAPH I



3-VARIABLE MULTIPLE CORRELATION
Graphical Approximation
$S_{1.23} = 2.87$
$R_{1.23} = .886 > R_{.01}$
Auxiliary Elevation
Heights in X_1 units
Edge view of plane 9, $\bar{X}_1$, 16 is edge view of approximated correlation surface.
8.63
34°
Plan
Elevation
$\bar{X}_3$
$\bar{X}_1$
$\bar{X}_2$

AUXILIARY ELEVATION through $\bar{\bar{X}}_3\bar{\bar{X}}_2$. The $\bar{X}_1$ values of 9, $\bar{\bar{X}}_1\bar{\bar{X}}_2$, and 16 are the heights required to complete the auxiliary view which will then show the edge of the oblique regression surface that is approximated by the plane 9, $\bar{\bar{X}}_1\bar{\bar{X}}_2$, 16. The use of points other than samples 9, 16, and the grand mean, will generally result in planes of slightly different directions so that a family of planes can be drawn. But, we are interested in the one plane that best represents the OVERALL data; hence, the use of averages is suggested as outlined above.

The balance of the sample values are projected into the auxiliary view. This should result in a RANDOM scatteration of the projected points about the estimated regression plane. We may check the randomness by inspection in some cases, or it may be necessary to plot a distribution of the measured deviations. The deviations are measured from the point to the plane coincident with the projection lines and perpendicular to the horizontal reference plane (HRP). When the values of these measured deviations are squared, summed, and divided by the number of samples, we have an estimate of the square of the standard error $S_{1.23}{}^2$. When it is close to a minimum with respect to the oblique regression plane, it will provide a satisfactory approximation of the problem solution. The value of the multiple correlation coefficient $R_{1.23}$ may be obtained directly from $S_{1.23}{}^2$ and $S_1{}^2$. In this case $S_1{}^2$ has been estimated from the range of the 18 sample values. (See Table I.) Between inspection of the graphs and testing $R_{1.23}$ for significance we are in a position at this point to make a decision regarding the analysis. The effect of the two independent variables is apparent when we compare the $S_{1.23}$ value with the S_1 value. In the auxiliary elevation, as shown on Graph I, the $3S_{1.23}$ limits are indicated as being 8.61 (in $\bar{X}_1$ units) above and below the oblique regression plane. The plane is referred to as point zero in order to correspond to the residual notation of $(\bar{X}_1-\hat{\bar{X}}_1) = z$, where z is the difference between the observed $\bar{X}_1$ value and the estimated regression plane. However, due to the plane's obliqueness (which occurs whenever there is correlation between the three variables), the deviations are not perpendicular to the correlation plane, but lie at some acute angle. In this example, the angle at which the residuals are projected onto the regression plane is measured at 54°. This is not necessarily the correct angle. This angle is a function of the $\bar{X}_1$, $\bar{X}_2$, and $\bar{X}_3$ scales used on the graph. Ignoring this in no way vitiates the determination of the $S_{1.23}$ value, but it does leave open the way for error when determining the perpendicular distances from the points to the regression plane. The reason for the discussion of this angle will become evident shortly. The scales should be standardized on the basis of their standard deviations; i.e., the unit distances on each scale will be inversely proportional to its standard deviation. For this example, the standardization is approximated, as noted in Table I, by making use of the ranges.

TABLE I

3-Variable Data*

Graphic Analysis

N = 18

n = 3 Sample No.	Dependent Quality Characteristic $\bar{X}_1$	Independent Variables $\bar{X}_2$	$\bar{X}_3$	Measured Deviations z
1	101.7	237	2.7	5.0
2	106.7	236	10.9	1.2
3	98.3	236	7.3	-3.5
4	95.3	234	5.9	-5.0
5	105.7	239	14.8	-3.7
6	105.7	206	8.6	1.2
7	106.0	213	10.0	0.2
8	107.7	203	6.0	5.7
9**	97.0	194	1.1	0.0
10	98.7	196	1.4	1.5
11	102.7	203	6.5	0.2
12	107.3	120	5.7	1.5
13	110.3	100	11.1	-2.2
14	109.7	106	12.5	-4.2
15	111.7	93	11.2	-1.5
16**	117.7	104	16.0	0.0
17	116.3	122	13.2	2.5
18	117.7	118	15.7	1.2
Sums	1916.2	3160	160.6	+3.1
$\bar{\bar{X}}$'s**	106.45	175.56	8.922	Refer to
Ranges	22.4	146	14.9	Graph I

1. Estimate of S_1^2:

$S_1 = R/d_2$

$S_1 = 22.4/3.64$

$S_1 = 6.2$

$S_1^2 = 38.4$

2. Estimate of $S_{1.23}{}^2$:

$S_{1.23}{}^2 = \text{Sum } z^2/N$

$S_{1.23}{}^2 = 8.24$

$S_{1.23} = 2.87$

$3S_{1.23} = 8.61$

3. Estimate of $R_{1.23}$:

$R_{1.23}^2 = 1 - S_{1.23}{}^2/S_1^2$

$R_{1.23}{}^2 = 1 - 8.24/38.4$

$R_{1.23} = .886$

* Coded

** Used for the graphical approximation

(S denotes standard deviation)

PRIME WESTERN ZINC METAL PRODUCT CONTROL

Prime Western Slab Zinc Lead Limit (1.6%) Control Factors

1. Secondary Furnace Feed
2. Sinter Feed Lead
   1. Calcine Lead
      1. Concentrate Feed Lead
   2. Cottrell Dust
3. Furnace Thermocouple Temperatures
   1. Preheat Temps.
   2. Fce. Charge Heights

Graph II shows an isometric view of the solid $\bar{X}_1\bar{X}_2\bar{X}_3$ with the oblique linear regression plane that is generated by the three varibles. This pictorial view serves the purpose of indicating a little more clearly how the negative effect of $\bar{X}_2$ and the positive effect of $\bar{X}_3$ together influence the dependent variable $\bar{X}_1$. By graphical solution the grand means $\bar{\bar{X}}_1\bar{\bar{X}}_2\bar{\bar{X}}_3$ and the points 9 and 16 lie on the linear regression plane. The dihedral angle (d) that the correlation plane makes with the horizontal base plane $\bar{X}_2\bar{X}_3$ is 36°. This may be determined from Graph I where the horizontal reference plane (HRP) intersects the approximated correlation surface at an angle which is the complement of 54°. We will refer to this picture again when we discuss the control chart analysis.

Referring once again to the auxiliary elevation shown on Graph I, the regression plane with the $3S_{1.23}$ limits suggests a time chart in terms of three variables. Accordingly, this auxiliary view is reproduced in Graph III and the residuals or measured deviations are plotted in TIME sequence. Also, from the angle of 54° we can orient the scale vertically so that we may think of the points as deviating perpendicularly with respect to the regression plane. When we do this, the limits become 0 $\pm$7.0. On this chart we observe $\bar{X}_1$ varying about a PLANE, the limits of which might be extended for use in plotting future data. The variations that we observe on this 3-variable chart are the "remains" after the effects of $\bar{X}_2$ and $\bar{X}_3$ have been removed, and these residuals might at this point be found to correlate with still another independent variable, $\bar{X}_4$. Also shown on Graph III is a frequency distribution of the measured deviations.

Tables II and III show, step by step, a convenient set of calculations for this 3-variable multiple correlation problem. The results of these calculations will provide the minimized solution which may be compared to the graphical solution. (The calculated $R_{1.23}$ = .910 as compared to .886, the graphical result) When the calculations are carried out, more detailed information may be had, such as the partition of the $\bar{X}_1$ variance. With regard to notation, the cyclic interchange of subscrips is inferred throughout. In Table III(b) the calculation of angle (d) is performed using the betas, the relation of which is indicated on Graph II.

The concept of multivariate regression as it relates to control chart analysis is best seen when we take the control chart for the $\bar{X}_1$ data as shown on Graph IV, and superimpose it on the oblique regression plane pictured in Graph II. It is obvious that we are viewing on the control chart a receding plane containing points both far and near as well as high and low with respect to our position in front of the chart. The near-far and high-low positions of the points are a function of the independent variables $\bar{X}_2$ and $\bar{X}_3$. Points 3, 4, and 9 below the lower

GRAPH II

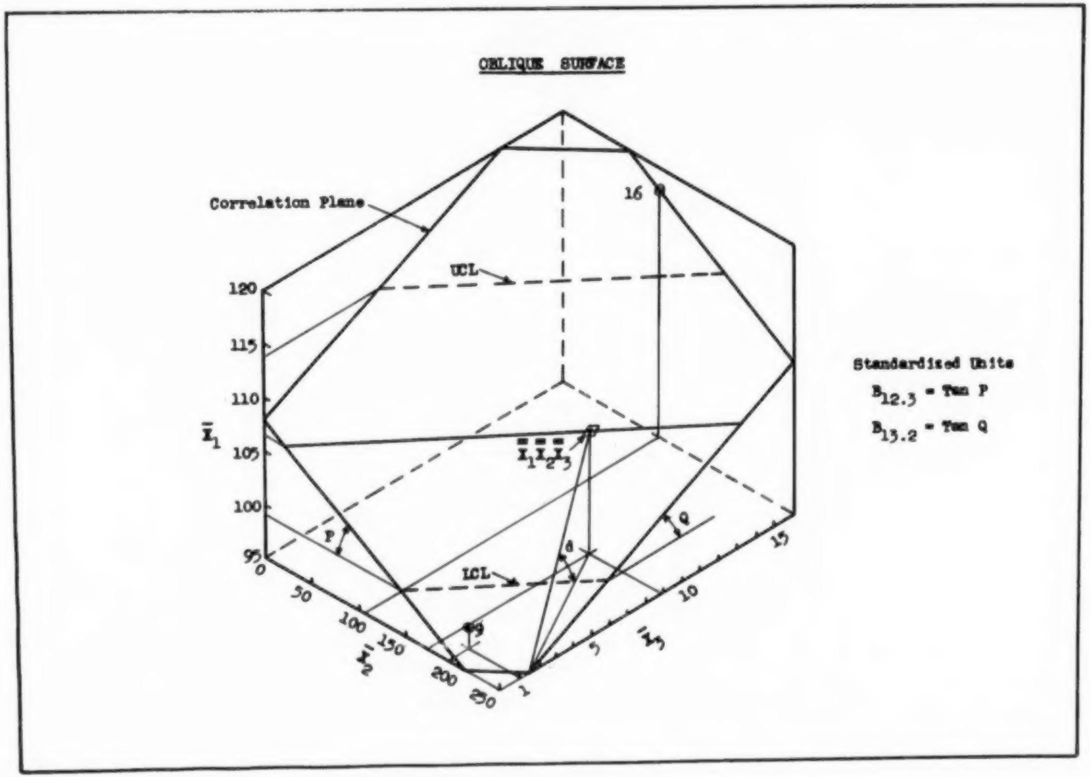

OBLIQUE SURFACE
Correlation Plane
UCL
LCL
16
9
$\bar{\bar{X}}_1 \bar{\bar{X}}_2 \bar{\bar{X}}_3$
$\bar{X}_1$
120
115
110
105
100
95
$\bar{X}_2$
0
50
100
150
200
250
$\bar{X}_3$
1
5
10
15
P
Q
Standardized Units
$B_{12.3}$ = Tan P
$B_{13.2}$ = Tan Q

TABLE IV

Control Chart For $\bar{X}_1$

Subgroup Size n = 3 Sample No.	Quality Characteristic (X_1) $\bar{X}_1$	Range R
1	101.7	8
2	106.7	10
3	98.3	6
4	95.3	1
5	105.7	12
6	105.7	9
7	106.0	8
8	107.7	5
9	97.0	8
10	98.7	10
11	102.7	9
12	107.3	6
13	110.3	7
14	109.7	3
15	111.7	9
16	117.7	6
17	116.3	1
18	117.7	9
Sum $\bar{X}$'s	1916.2	-
Sum R's	-	127
$\bar{\bar{X}}_1$	106.45	-
$\bar{R}$	-	7.05

$3S_{\bar{X}_1} = A_2\bar{R} = 7.22$

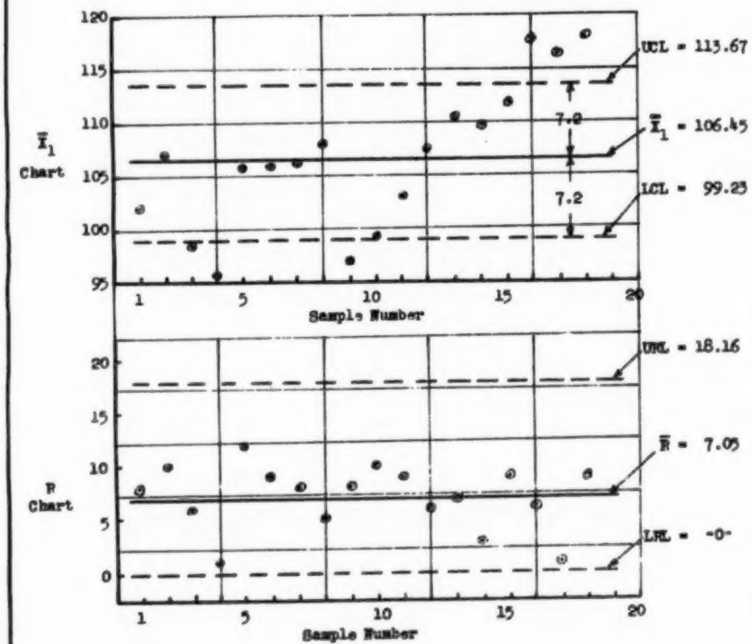
$S_{\bar{X}_1} = 2.41$

Control Chart Limits

	U C L	L C L
$\bar{X}_1$	113.67	99.23
R	18.16	-0-

Refer to Graph IV

GRAPH IV



CONTROL CHART

n = 3

$3\sigma_{\bar{X}_1} = 7.22 \quad (A_2\bar{R})$

control limit are at: high $\overline{X}_2$ - low $\overline{X}_3$; and points 16, 17, and 18 above the upper control limit are at: low $\overline{X}_2$ - high $\overline{X}_3$. While sample points like 9 and 16 show out of control on Graph IV, actually both are on or near the oblique plane and fit the 3-variable function very well. But, with respect to specifications for $\overline{X}_1$ as they might coincide with the control chart limits, points 9 and 16 ARE OUT of "spec", and a graph like the 3-variable chart (Graph III) will not tell us when $\overline{X}_1$ has exceeded those specifications. The power of the control chart technique is evident when we compare the $3S_{\overline{X}_1}$ limits with the $3S_{1.23}$ limits.

This study of multivariate regression and its relation to control chart technique brings up the question--when could expanded analyses such as this be used to best advantage? As a general rule, most processes may be thought of in terms of two major parts--the physicochemical and the mechanical. Emphasis placed on correlation, analysis of variance, and similar techniques seems to have certain direct advantages when solving production and quality problems prior to the mechanical part of the process. A dividing line at the finished pig, ingot, or slab is suggested as being the point beyond which control chart technique becomes more applicable. In either case, however, it is the exception rather than the rule to find a dependent variable, over a given period of time, free of "outside" effects. Often certain results in a process are conditioned by variables several steps removed from the point under study. These effects can be transmitted either directly or indirectly by means of interactions with other major process variables. What may be a dependent variate in one part of a process often becomes an independent variable with respect to the next step in the process. Simplified multivariate analysis, it would then appear, has at least this one desirable feature: it enables the Quality Control Engineer to quickly determine variables that might be causing a control chart to show excessive variations. Determination of the type of problem and consequently the statistical treatment to be used can be important analytical decisions.

References:

"Elementary Statistics and Applications", Smith & Duncan, Volume I, McGraw-Hill, 1944.

"Control Chart Method", American Standards Association, New York, N. Y.

GRAPH III

3-VARIABLE CHART

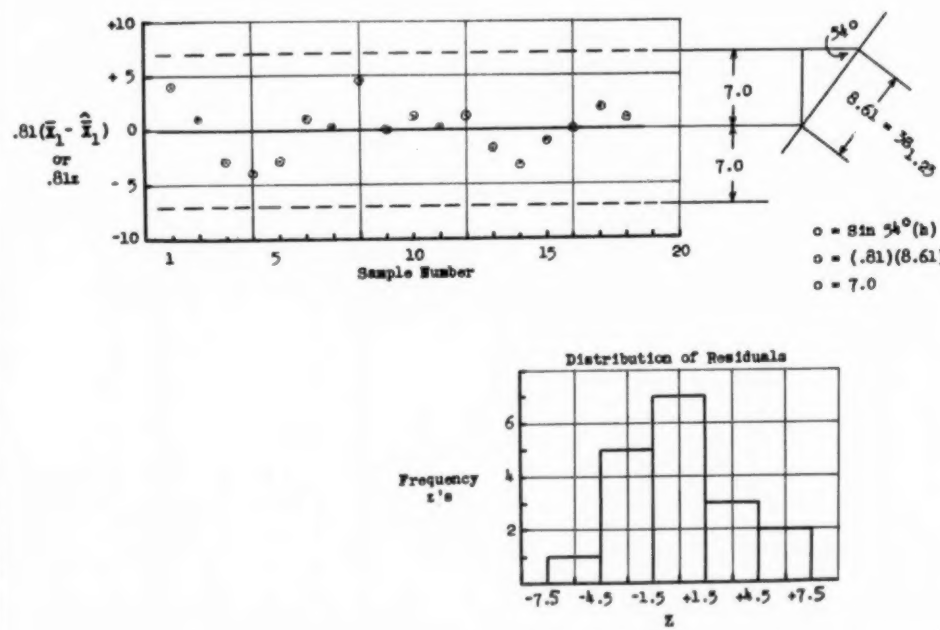


TABLE II

3-Variable Calculations
(Data in Table I)

Sums of Squares:

$\text{Sum } x_1^2 = \text{Sum } \bar{X}_1^2 - \bar{\bar{X}}_1 \text{ Sum } \bar{X}_1$

$\text{Sum } x_1^2 = 783.5$	$\text{Sum } x_2^2 = 54866.$	$\text{Sum } x_3^2 = 369.8$
$s_1^2 = 43.53$	$s_2^2 = 3048.$	$s_3^2 = 20.54$
$s_1 = 6.60$	$s_2 = 55.2$	$s_3 = 4.53$

$\text{Sum } x_1x_2 = \text{Sum } \bar{X}_1\bar{X}_2 - \bar{\bar{X}}_1 \text{ Sum } \bar{X}_2$

$\text{Sum } x_1x_2 = -4887.3$	$\text{Sum } x_1x_3 = +437.2$	$\text{Sum } x_2x_3 = -2126.8$

Simple Correlation Coefficients:

$r_{12} = \text{Sum } x_1x_2/(\text{Sum } x_1^2 \text{ Sum } x_2^2)^{1/2}$

	r	$1-r^2$
$r_{12} =$	-.7454	.4444
r_{13}	+.8122	.3403
r_{23}	-.4722	.7770

Partial Correlation Coefficients:

$r_{12.3} = (r_{12} - r_{13}\, r_{23})/(1-r_{13}^2 \cdot 1-r_{23}^2)^{1/2}$

$r_{12.3} = -.7038$

$r_{13.2} = +.7832$

Standardized Multiple Regression Coefficients:

$B_{12.3} = (r_{12} - r_{13}\, r_{23})/1 - r_{23}^2$

$B_{12.3} = -.4658$

$B_{13.2} = +.5923$

TABLE II - Continued

Residual Error:

$$S_{1.23}^2 = S_1^2(1-r_{12}^2)(1-r_{13.2}^2)$$

$$S_{1.23}^2 = 7.47$$

$$S_{1.23} = 2.73$$

$$3\,S_{1.23} = 8.19$$

Multiple Correlation Coefficient:

$$R_{1.23}^2 = 1 - S_{1.23}^2/S_1^2$$

$$R_{1.23}^2 = .8283$$

$$R_{1.23} = .9101 > R_{.01}\ \text{d.f.} = 15$$

Multiple Regression Coefficients:. (in original Units)

$$b_{12.3} = B_{12.3}S_1/S_2$$

$$b_{12.3} = -.05567 \qquad b_{13.2} = +.8621$$

Multiple Regression Equation Constant:

$$a_{1.23} = \bar{\bar{X}}_1 - b_{12.3}\,\bar{\bar{X}}_2 - b_{13.2}\,\bar{\bar{X}}_3$$

$$a_{1.23} = +108.54$$

Multiple Regression Equation: (for 3-Variates)

$$\hat{\bar{X}}_1 = +108.54 - .05567(\bar{X}_2) + .8621(\bar{X}_3)$$

Partition of $\bar{X}_1$ Variance:

$$100\,(r_{12}^2 + r_{13.2}^2\,(1-r_{12}^2) + S_{1.23}^2/S_1^2) = 100 \text{ per cent}$$

Per cent $\bar{X}_1$ Variance	Due To	Calculation
55.6	$\bar{X}_2$ effect	r_{12}^2
27.2	Addition of $\bar{X}_3$	$r_{13.2}^2\,(1-r_{12}^2)$
82.8		$R_{1.23}^2$
17.2	Residual error	$S_{1.23}^2/S_1^2$
100.0%		Total

TABLE III

(a)

Calculation of Residuals from Multiple Regression Equation:

$\hat{X}_1 = +108.54 - .05567(\bar{X}_2) + .8621(\bar{X}_3)$

Calculated Residuals

Sample No.	Values $\hat{X}_1$	$(\bar{X}_1-\hat{X}_1)$ z
1	97.7	4.0
2	104.8	1.9
3	101.7	-3.4
4	100.6	-5.3
5	108.0	-2.3
6	104.5	1.2
7	105.3	0.7
8	102.4	5.3
9	98.7	-1.7
10	98.8	-0.1
11	102.8	-0.1
12	106.8	0.5
13	112.6	-2.3
14	113.4	-3.7
15	113.0	-1.3
16	116.5	1.2
17	113.1	3.2
18	115.5	2.2
Sums	1916.2	0.0

$S_{1.23}{}^2 = \text{Sum } z^2/N$

$S_{1.23}{}^2 = 7.49$

$S_{1.23} = 2.73$

(b)

Calculation of Dihedral Angle (d) and Perpendicular Distance to Regression Plane:

Angle between Multiple Regression Plane and Base Plane $\bar{X}_2\bar{X}_3$.

$\text{Tan (d)} = (\text{Tan}^2 \text{ P} + \text{Tan}^2 \text{ Q})^{1/2}$

$\text{Tan (d)} = (B_{12.3}{}^2 + B_{13.2}{}^2)^{1/2}$

$\text{Tan (d)} = .7535$

Angle (d) = 37° in terms of standardized units based on the standard deviations

Angle the Vertical Projections of the Residuals Make With the Regression Plane

90° - Angle (d) = 90 - 37 = 53°

Perpendicular Distance of Residuals to Regression Plane

Sine 53° = o/h, where h = z

O = (Sine 53°)(z)

Perpendicular Deviations are

.7986 (z)

Refer to Graphs II & III

STATISTICAL DESIGN IN ELECTRONICS PRODUCTION-LINE EXPERIMENTATION

Frank Caplan, Jr.
General Electric Company

In the manufacture of aluminized television tubes, one of the most important operations is the depositing of a nitrocellulose lacquer film to provide a smooth base for the aluminum mirror. There are several methods for performing this operation, but the one on which the following study was made was a flotation filming process. In this process a quantity of water is added to the bulb after screening; a quantity of nitrocellulose lacquer is then deposited on top of the water and permitted to spread. Afterwards the water is poured out from underneath the film which then adheres to the back of the screen. This type of process has always been characterized by a high average shrinkage rate with large fluctuations between very good results and very high shrinkage.

At the time of the initiation of the study, the shrinkage rate was relatively high and the item contributing the most to this high shrinkage rate was "holes in the film". Under these circumstances it was felt that a statistical investigation would be very useful. Normal engineering and production approaches had been found incapable of permanently solving the problem. This was because of the existence of a large number of possible causes for holes, the relative importance of which was not known. In addition, it was felt that it was quite possible that there were significant interactions on several of these sources of variation, thus complicating the problem a great deal. The adoption of a statistical approach to the solution to this problem would result in the greatest economy of experimentation, and also should eliminate the bias of persons who had been working on the problem, "pet ideas" and things of that nature.

It was felt that the experimentation should be performed in the production operation, since duplication of all the conditions of the production processes would be almost impossible on a laboratory scale. The first step in the experimental procedure was to institute a preliminary survey. This was followed through in an orderly fashion, starting with observation of the procedures in use, and checking these against established published methods to determine whether there were any important variations from practice as previously written up. Then discussions were held with the production supervision, the engineers involved, planners, the operators, machine attendants, the maintenance people and everyone else who might have any ideas concerning the possible sources of the difficulty. As a result of these discussions the list of all factors which might possibly cause variation in the opinion of any of the people who were talked to reached a total of forty. Some of these were obvious from the beginning, and some were not. A few of them were items which could not be controlled by any experiment or by anyone working on the process, except possibly by very expensive installations of equipment. Most of these, however, could be measured even though they could not be controlled. Other factors could be controlled but not measured. And a third group could be both measured and controlled. It was from this third group that the factors were to be selected for study.

The first group chosen were those which had been mentioned as possible sources of holes by the greatest number of people interviewed.

At this time control charts and special shrinkage records were set up to give the picture over the entire course of the investigation showing what improvements were made. It was decided that two experiments could be run simultaneously, each comparing three levels of four variables. In order to have a reasonable estimate of error, the resultant Graeco-Latin square design was replicated four times. The variable measured to give the estimate of resultant quality was an arbitrary score based upon the diameter of the hole and the quantity of holes of each size recorded.

In the first experiment all four sources of variation were determined to be significant. However, one was significant far beyond the limits of the table, one significant at the 1/10th of 1% level, and the other two significant at the 5% level. This is illustrated in Table 1. In each case we were able to reject one level and have relatively little preference between the other two. This gave us an area in which we could work without too much variation in the end result.

The second experiment, run at the same time as the first, yielded only one significant source of variation, as noted in Table 2. In this one also there was a rejection of one level with relatively little choice between the two remaining.

The third experiment was a Latin square design with just three factors, each at three levels, one of them repeated from Experiment I to verify some rather unexpected results which were obtained in this experiment. Table 3 shows the result of this analysis. All three factors were significant, one at the 1% level, the others beyond the limits of the table. This clarified the effect of the variable carried over from the first experiment and gave us a choice of one particular level of each of the three factors involved.

At the same time we ran Experiment IV, which was a full factorial design of two variables, each at three levels. The full design was chosen because of the high degree of possibility of the existence of a significant interaction between the two factors chosen. As the analysis in Table 4 shows, however, this interaction did not prove to be significant, and only one of the main effects was a significant source of variation. This gave us a definite choice of one level of this factor and a rejection of the other two.

Experiments III and IV were run under controlled conditions at the best levels of the factors that had been studied in Experiments I and II. This resulted in such a great improvement in the holes found during this investigation that it made analysis rather difficult. Therefore, at this point it was decided to issue a report informing the factory of the advances made to date and permitting them to incorporate these improvements in their process. This was done with a marked resultant reduction in shrinkage attributable to holes.

Experiments V and VI were then run, partially to verify the results found previously and partially to extend the areas investigated on certain significant variables. This permitted us to optimize the results giving the best possible results obtainable with the materials we were using. Experiment V was a three by three Latin square yielding us no new information on holes. This is illustrated in Table 5.

Table 6 also shows no new information concerning holes; but anticipating this, we designed it to give significant information about some

of the other defects noted in the filming operation. This was a four by four Latin square, again intended to permit us to optimize the process. This was quite successful and yielded an optimized process which has proved to be quite workable. Holes have been reduced as a source of shrinkage by more than 50%, even though some of the expensive work involving large equipment installations has not yet been undertaken. The stability of the overall process has also been greatly improved.

Source	Sum of Squares	df	Mean Square	"F" Ratio
A	31,939	2	15,970	4.97*
D	132,739	2	66,370	20.67****
G	36,141	2	18,070	5.63*
K	68,091	2	34,046	10.60***
Residual	57,803	18	3,211	
Total	326,713	26		

Table 1. Summary of Variance - Experiment I

Source	Sum of Squares	df	Mean Square	"F" Ratio
N	4,006	2	2,003	5.08*
Q	367	2	184	
T	434	2	217	
W	747	2	374	
Residual	10,644	27	394	
Total	16,198	35		

Table 2. Summary of Variance - Experiment II

Source	Sum of Squares	df	Mean Square	"F" Ratio
α	211,739	2	105,870	33.22****
δ	122,980	2	61,490	19.29****
G	53,100	2	26,550	8.33**
Residual	89,235	28	3,187	
Total	477,054	34		

Table 3. Summary of Variance - Experiment III

Source	Sum of Squares	df	Mean Square	"F" Ratio
Θ	5,745	2	2,872	
λ	198,690	2	99,345	22.83****
$\Theta \times \lambda$	26,055	4	6,514	1.68
Residual	69,672	18	3,871 (4351)	
Total	300,162	26		

Table 4. Summary of Variance - Experiment IV

Source	Sum of Squares	df	Mean Square	"F" Ratio
λ'	505	2	252	1.98
δ	93	2	46	
θ'	510	2	255	2.00
Residual	3,700	29	127.5	
Total	4,808	35		

Table 5. Summary of Variance - Experiment V

Source	Sum of Squares	df	Mean Square	"F" Ratio
A'	2,639	3	879	
π'	989	3	329	
α'	7,477	3	2,492	2.54 (9%)
Residual	37,228	38	980	
Total	48,333	47		
	for Holes			
A'	51	3	17	4.15**
π'	182	3	60.7	14.80****
α'	175	3	58.3	14.22****
Residual	157	38	4.1	
Total	565	47		
	for Other Defects			

Table 6. Summaries of Variance - Experiment VI

QUALITY CONTROL LOOKS AT THE FINANCIAL STATEMENT

J. Leslie Lenton
American Machine & Foundry Company

Politically and industrially, during the last few years, everyone has been hearing about the high cost of living. Let us proceed from this point of common knowledge to concern ourselves with the cost of high living, rather, - - - the cost of high industrial living, as it is reflected in the financial statement.

No doubt your company publishes a financial statement, and, providing this is the case, have you taken time to read it, and, if you did, do you understand it? Could you pinpoint the contribution that your quality control department has made to your company's fiscal bill of health?

Do you realize that it may be more vital to your future planning to interpret the statement in terms of what is omitted than in sum totals as they are presented?

In our management courses at American Machine & Foundry, it has been our policy first, to select the right men for training, and second, to offer these men the right training.

Most things for our good must be, by their very nature, direct and with a purpose. And so today, if my remarks are to be of any value to you, they must be direct, -- plain-spoken, and with a purpose.

You are interested in your pay check. Your solvency depends on it. Men of your training and profession must be equally interested in the financial report which testifies to the solvency of your company.

The financial report is an instrument common to all industries, whether large or small. Let us begin with a smaller picture which will be easier for us to discuss and understand. I would like to begin with you. You conduct most of your personal and domestic bill-paying by check. Very little actual cash passes through your hands. It is conceivable that you may handle less than ten or fifteen percent of your pay check in cash. Most of us find it very difficult to comprehend the magnitude of business which is transacted every day without the actual transfer of one silver dollar.

The day when the paymaster filled little envelopes with crisp new bills and bright half-dollars is gone. We are living in a period of shifting economy. The implication of this fact to all of us is somewhat like diseases, which do come upon us very gradually and without our realizing it. This necessarily then calls for a "New Look". Our President has used this very term.

In order to take a "New Look" at our economy, we must be acutely aware of what is going on around us. Spiritually ... Emotionally ... Physically ... Financially.

Several years ago, quality control was the "New Look". Every industry is quite generally grateful for what the Quality Control Society has done and will continue to do in developing a most dynamic force aiding and promoting cost reduction and improving the quality of

a vast variety of products.

At American Machine & Foundry, quality control has become a strong staff function. The quality control department has taken a "New Look" into the financial aspects of business management, which is why I am here. We work as a team, planning, controlling, accounting, crediting, debiting, -- acting jointly. We think in terms of the "New Look", of how all our efforts will be reflected in our company's financial statement.

What about the "Old Look"?

Fourteen hundred years before Christ the rulers of Egypt were entombed with fabulous wealth, so they could pay their way into eternity.

The Gospel according to Matthew, tells of money changers in the temple itself.

Marco Polo's letters from the Orient, written in the fourteenth century, described busy market-places where pieces of gold changed hands for jewelry, fabrics, and other products of the East. Purchase ---- Payment.

In 1596 Shakespeare wrote "The Merchant of Venice", a tale of a man who loaned money and asked that it be repaid with interest. Borrow --- Repay, Profit --- Loss, Asset --- Liability.

In former times, vast sums were not mere words. They were quantities of gold, or gems, or silver, or land. They could be touched, seen, counted, measured. A man had but to total his wealth to determine whether or not he was solvent.

In the early days of our own country, barter was common practice. The harvest of a field of wheat was valued in terms of other things, cattle, pigs, plow points. Potential assets were tangible. A man's financial status was almost visible. A small business man could borrow on his word. His signature represented his financial statement, and acted as his collateral.

After World War I we became accustomed to the sound of the word MILLIONS. Some of us did not always know what this meant, or just how big a pile of gold dust it would take to make a million dollars. We comprehended it better when we were told that a million dollars equalled a string of ten dollar bills reaching from New York to Sacramento, California. But some of us had never been to California, so even that did not quite register.

World War II came along, and the word MILLIONS became as out-of-date as the Liberty engine. We had a new word - BILLIONS. The public debt --- now 278 BILLIONS. What does it mean? What does it mean? It means that we can no longer count our wealth in terms of cold, hard cash.

We have with us the income tax, the social security tax, deductions of all kinds, and they are pushing the old form of bookkeeping into obsolescence along with the quill, the ink pot, and the four-row adding machine. We have arrived at the time when the manufacturers of

calculating machines and electronic tabulation systems hesitate to sell anything less than twelve key-bank calculators. High time we took a "New Look" at our economy.

The bookkeepers are not the only ones who have seen remarkable changes. Within the past year, there have been notable advances in the fields of science and legislation. Let us review some of them briefly:

February 16, 1953 ----- Removal of control on wages and salaries.

March 17th ------------ Price controls end.

May 25th -------------- First atomic cannon fired in Nevada.

June 30th ------------- End of fiscal year. Federal deficit reaches new high. Controlled material plan ends. Amen.

August 18th ----------- New Government Power Policy removes responsibility for local power needs from industry and local government.

October 28th ---------- Bureau of Statistics reports cost of living index at all-time high.

December 20th --------- Prediction that no more piston-type aircraft engines will be purchased by Government after 1954.

December 29th --------- Prediction that President Eisenhower's State of the Union Message will be reflected in financial statement of the nation's industry. And-don't-you-think-it-won't.

January 27, 1954 ------ The first atomic battery becomes a reality.

There must be a common denominator. And, of course, there is. The common denominator is always money. Less government control means more free enterprise. Free enterprise means that industry must learn to stand on its own two feet.

Let us consider your attitude toward the daily stock market reports published in your newspaper. As long as you have no money invested in stocks, or securities, you are quite likely to skip the financial section entirely and turn to the comics. But as soon as you have accumulated enough cash to make an investment, as soon as you have become a stockholder, you discover the fine print on the financial page. The stockholder, and now this means you, has developed a new interest. He wants to interpret the stock reports in the light of possible dividends. He thinks, now, in different terms.

When you have a cash investment in some industrial enterprise you either "point with pride", or "view with alarm" the daily rises or dips in stock quotations. These shifting figures reflect business

activity, and you are likely to conclude that business is either good or not so good, as the figures go up or down. You are interested in earning dividends. Your money must work!

Who does declare dividends? This may, in most instances, be performed by the Board of Directors. If it is done by the Board of Directors, then how do they know when to vote a dividend? Do they have a mystic power or magic means that is known only to a few? .... Unfortunately, they do not .... They decide when to vote a dividend on the basis of reports prepared by the financial division of their organization. What reports, then, must the Board review in order that their judgments may be sound?

There are three reports which are basic:

(1) The Balance Sheet.

(2) The Profit and Loss Statement.

(3) The Manufacturing Expense Statement.

These three reports, together with other supporting schedules and appendices, furnish the material figures on which the judgment of the Board of Directors must be based. These reports are similar, no matter what industry is under discussion, but to be certain that we are not confusing our terminology, let us state briefly the purposes of these three reports:

First, the Balance Sheet. This is intended to show the financial condition of the business at a particular time, usually at the end of a designated period ... calendar year ... or fiscal year. The Balance Sheet alone is merely an expression of the Auditors' opinion of the financial condition of a business at a given time. Much must be read into this report and its notes.

Second, the Profit and Loss Statement. This is the economic summary or income statement. It summarizes the changes which have taken place since the beginning of the period referred to in the Balance Sheet, and also shows how these changes affect the business, favorably or unfavorably. In other words, it indicates how the profit or loss came about. In the larger corporations, both the Balance Sheet and the Profit and Loss Statements will be prepared in detail as well as briefed in a more condensed form.

Third, the Manufacturing Expense Statement. This report indicates those items which account for all the expenses incurred throughout the manufacturing process. You will readily agree that the Manufacturing Expense Statement influences both the Balance Sheet and the Profit and Loss report of the company.

The Annual Statement then, is the sum total of the many supporting reports and documents which have been accumulated during the fiscal period. The best modern management practice demands the constant flow of this live information through the executive office while it is timely and meaningful. It is of prime importance that reports designed by the accounting function be supplied promptly and regularly. I need not remind you quality control men that the value of a report depends entirely on the time elapsed between the event and the report. In other

words, the value varies inversely with the lapse of time.

You know and I know that no accounting system, properly designed for the use of management, places the entire emphasis on financial statements as its primary or final goal. But we must learn to think in terms of money. Our efforts, good or bad, are reflected in the financial reports of the companies for which we work.

The quality man is often concerned with standards of performance. We have said little about this. But, a quality manager's standard of performance is related to the figures which he is in a position to influence. He is in a position to influence the daily reports which flow through manufacturing, accounting, to management, and then to the Board of Directors.

A good quality control manager speaks two languages. To the technical and production personnel ... he speaks of <u>things</u>. To the executive management, he must talk of <u>dollars</u>. But <u>I tell</u> you no quality control manager will ever be a success unless he is able to think in terms of <u>dollars</u> ... while he is talking about things.

The quality control manager must concern himself with further reports which contribute to the formation of the financial statement and, as you know, such type reports are necessary and numerous.

It should be understood that not all of these reports which contribute to financial-statement-make-up can be, or will be influenced by the quality control function.

Let us examine those reports which you <u>can</u> influence.

<u>Net Profit and Loss</u>

It is obvious that, if you neglect to <u>think</u> and <u>act</u> in terms of dollar operations, you directly influence the profit figure ..... downward!

<u>Indirect Manufacturing Expense</u>

Quality control is a part of this expense. Therefore, any production dollar saving techniques appear in the profit or loss picture directly.

<u>New Business</u>

We all know that consumer buying is a habit. New business comes from old satisfied customers, or new customers who think they will be satisfied because of your product quality reputation. You must have optimum quality at optimum prices. You must effect a compromise between inferior goods and solid gold Cadillacs.

<u>Job Efficiency</u>

There are multiple uses of this report. How does quality control activity relate to the Balance Sheet via the Job Efficiency record? Say, your plants ran 20% defective in 1952. By diligent quality control effort it was reduced to only 5% in 1953. I mean, all materials, supplies, and inventories were 5% defective. You may feel that you have

accomplished great strides in process quality improvement. No doubt you have. Nice work. But must we look further?

From top management's level, the Balance Sheet would show another point of view. Looking on the asset side, you will find under Current Assets, the value of raw materials and inventories. In a large multi-plant corporation, it is not uncommon to find this figure upwards of several millions of dollars.

For instance, at AMF, this item for 1952 represented a very large sum. Let us say that 5% of it would amount to about one million dollars. You can see that such a rejection loss, compared with the net profit for the year, would be most startling and absolutely intolerable.

So you can see how your quality control activity is reflected upward to the asset side of the financial statement.

As you become aware of the importance of the financial statement and begin to resolve its many implications, you will see why modern management considers its quality control group as an asset instead of a liability.

You will also learn that there are several companies which operate on a margin of as little as a 3% net profit. This usually occurs in large and heavy industries. Sometimes only a small percentage of your raw materials or work in process may equal, or even exceed, your company's gross dividend for the year.

There could be no more direct approach to influencing your company's financial statement than by acting on the principles I have outlined: "The effectiveness of any quality control department is directly reflected by the favorable change in the dollar figures which it is in a position to influence".

If I have convinced you that in a single plant operation, quality control should look at the financial statement, then it must be even more obvious that in a multi-plant organization, quality control should assume its responsibilities as a modern management financial tool.

In summary, I would like to state a seven-point program:

(1) Quality control people are too often bound by their own control limits. Go on the alert, look around you, try to learn the language of your executive management, especially the planning and control functions connected with dollar activities.

(2) Quality control shares in the making of your company's financial statement, whether it wishes to or not, and its function will, in a large measure, decide whether the statement is favorable or unfavorable.

(3) Those engaged in quality control activities should learn accounting principles and practice. This will aid them to better understand the influence that they do create on their company's financial statement.

(4) Study your own financial statement. Ask yourself ... "How have we contributed to the Asset side? If you do not know, find out.

(5) Go on the offensive! Ask your accounting department for assistance in obtaining current records. Some of the figures may shock you. If you take action, this is good. If you do not, somebody else <u>might</u>!

(6) If you are a subsidiary and purchase from a co-subsidiary, set up uniform standards. Eliminate duplicate inspection. We do. It pays dividends.

(7) You may have many plants in your organization. You will have but one consolidated financial statement - the annual report.

Like engineering standards of practice, quality control should be set up and administrated by those who thoroughly understand the function of a well-organized and disciplined quality control department.

If our discussion with you today has provoked a challenge, then our mission will have been fulfilled. Should this be the case, from this day forward, you will recognize your financial statement as a mirror to you. Such a mirror would reflect only what is put into it -- in other words, what it sees.

You will --- and you should --- control the image this mirror reflects.

STATISTICAL METHODS FOR APPRAISING PUBLIC UTILITY PROPERTY

E. T. Magruder
The Chesapeake and Potomac Telephone Companies

1.0 PURPOSE

In common with other business, a major task of the telephone industry over the past several years has been that of adjusting its rates to meet rising costs of operation. All such rate changes are subject to approval by a Public Utility Commission, or similar public authority. Among the many factors of public utility operation that are considered in detail by a Commission in a rate case proceeding is the value of its property.

One aspect of public utility property valuation that is usually important concerns the question, "What is its present value?". Even a question this simple is often difficult to answer satisfactorily. The property of a telephone company serving an entire state comprises a vast array of elements. Familiar property items like buildings, poles, cables, wires and telephones are exceeded in number and complexity by the switches, relays, switchboards and their appurtenances that make possible the network of communications.

Value at a given time is obviously a function of the physical condition of this myriad of parts. Measurement of physical condition has, of course, been an essential factor in utility property valuation since the inception of rate regulation by public authority. Inspection of a sample of 10% of the property by qualified experts was the usual procedure prior to the introduction of statistical sampling. Measurement of the physical condition of a tenth of the property was unduly costly and time consuming. Although a cross-section of this size was no doubt a reliable indicator of the average physical condition of the property, there was no way to compute its precision as selection of the items was mainly based on judgment. In addition, the sheer size of the job made it difficult - if not impossible - to select and train the required number of people with good assurance that their work was of uniform quality.

Under the guidance of sampling experts such as Prof. John W. Tukey and Dr. W. Edwards Deming several Bell System Companies have used statistical sampling to measure physical condition of property for use in rate case proceedings. This procedure has been to the advantage of both the Public Utility Commissions and the telephone companies. By this means, physical condition data of demonstrable precision was available for the first time - and, of like importance, it could be produced quickly and at relatively low cost. For example, a statistical sample covering about 30,000 items of property in a typical company established the average physical condition to a precision of ± 1/3% at the 3-sigma confidence level. A period of 4 months spanned the entire job at a cost of about 75 cents per $1,000 of total property. In fact, knowledge gained in the initial sample about the distribution of items by physical condition within the various classes of property makes possible even smaller samples in future studies. For this same company a condition study was made in late 1953 at a cost of about 10 cents per $1,000 of property. This smaller sample was geared to a precision range of ± 1% at the 2-sigma confidence level - and was mainly to test for any significant changes in average physical condition over the intervening five years since the more detailed study.

It is the purpose of this paper to give an outline of the sampling procedure that has been used to measure the average physical condition of telephone property. In brief, ten sub-samples were selected from each major class of property, using a systematic sampling interval from a random starting point. Size of sample from each property - class - stratum takes into account the dual factors of variability in physical condition and relative value of the property by strata.

There are three phases to be considered, viz:

1. Sample design and selection of property items to be inspected.
2. Inspection procedure.
3. Compilation of results, including precision of measurement.

2.0 SAMPLE DESIGN

2.1 Outline of Procedure

There are five distinct phases in the design of a sampling survey, viz:

1. A determination of the objective precision range and the specified level of assurance that can be placed in the survey results.
2. A definition of each of the classes of property to be surveyed.
3. A determination of the sample size for each property class.
4. A specified method for selecting the sample items in each property class, together with the computing methods applicable in determining the findings and the precision range thereof.
5. The preparation of instructions for the field inspectors, design of forms for recording the inspection results, selection and training of inspectors and their supervisors, and development of an administrative procedure.

In the case of property surveys designed to be a part of rate case proceedings it is essential that the Company's expert witness on valuation be the administrator of the entire survey. Considerable technical assistance will, of course, be required, especially during the design of the survey. There is a substantial advantage in retaining an outside consultant, whose knowledge and proficiency in the science of sampling is generally recognized by experts in the field, to participate in the survey and to give testimony on the survey in rate case proceedings.

2.2 Precision Range

The precision ranges for individual property classes and for the property as a whole have an important bearing on the acceptability of the survey and also on its cost. For use in property valuation, a precision range of about ± 1.0% for the property as a whole - and in which an extremely high assurance can be placed - can be attained. Rigorous results of this degree are relatively expensive, however. It should be borne in mind that a precision range of ± 2.0% requires only 1/4th as large a sample for the same level of assurance as does a ± 1.0% tolerance. On balance, a property valuation as determined by sampling which almost certainly differs by less than 2%, or even 4%, from the exact value may be fully as useful as a design that insures a 1% precision range. While the 1% precision range has been used in several rate case proceedings, there appears to be a reasonable doubt that the added costs justify the small gain in precision.

An empirical scale that has proven generally satisfactory for sample size determination is as follows:

1. Rank the major property accounts in order of size(i.e. investment).
2. Assign precision ranges of ± 2%, ± 3%, and ± 4% (at the 99.5% assurance level) to the accounts in the upper, middle and lower thirds, respectively.
3. Determine the size of sample for each of the component classes of property within the several accounts which, in the aggregate, will provide the ranges specified.

Figure 1 shows the relation between the objective precision ranges and those obtained during an actual survey.

In the overall, the precision range came out to be almost exactly on the objective. While most of the 16 accounts showed a precision somewhat better than called for, this is explained by either over-esti-

Figure 1 - PRECISION RANGES, OBJECTIVE VS ACTUAL

i	Account (In Size Order)	Decimal Proportion of Total (W_i)	Precision Range (99.5% Assurance Level) (r_i)	
			Objective	Actual
I	Upper Third		±%	±%
1	Aerial Cable	.1770	2.0	2.19
2	C.O. Equipment (S x S)	.1229	"	3.11
3	Pole Lines	.1050	"	1.52
4	C.O. Equipment (Manual)	.0970	"	1.20
5	U.G. Cable	.0876	"	1.85
6	U.G. Conduit	.0638	"	1.56
II	Middle Third			
7	Station Installation	.0558	3.0	1.80
8	Station Apparatus	.0557	"	1.03
9	Private Branch Exchange	.0501	"	5.51
10	Aerial Wire	.0446	"	1.38
11	C.O.E. (Toll Exch. Serv.)	.0386	"	1.50
III	Lower Third			
12	C.O.E. (X-bar)	.0295	4.0	0.13
13	Drop & Block Wires	.0272	"	2.14
14	C.O.E. (Panel)	.0237	"	1.09
15	Motor Vehicles	.0154	"	3.54
16	Booths & Spec. Fittings	.0062	"	6.70
	Total	1.0000	±.68%	±.69%

The Equation for combining the precision ranges for the individual accounts into an overall precision range for the entire property is:

$$R\ (\text{overall}) = \sqrt{\sum (Wr)_i^{\,2}}$$

mating the variability or by an actual over-sampling of certain items that results by the inclusion of several property classes in one "sampling unit".

2.3 Sampling Units

Utility property is made up of a wide array of distinct classes of property. It is possible to sample each of these classes independently, but this is unnecessary. It is more economical to select a "pole-location", for example, as a "Sampling Unit". In addition to the pole itself, there are several other distinct classes of property at a pole location, viz: crossarms, anchors, guys, aerial cable, copper line wire, iron and steel line wire, cable terminals and drop wires. Hence a sample of pole-locations likewise provides a sample of the aerial property classes carried on poles.

A convenient sampling unit for underground plant is a manhole, as the cable, conduit and manhole can be inspected at each sample location.

In the case of Central Office equipment, the predominant class of equipment (in terms of the investment) can be used as the sampling unit. The step-by-step switch, for example, would be the sampling unit in offices of this type. After determining the number of units to sample of the predominant class, the sample sizes for the remaining classes can be roughed in by proportioning. That is, if Class B contains half as large an investment as Class A (the predominant class), then the sample for Class B should be about half that for Class A. This procedure is only a general guide, of course, as the known (or estimated) degree of variability in physical condition of the subsidiary classes forms a basis for deciding appropriate sizes of samples.

In brief, the choice of a sampling unit should be based on two considerations, viz:

1. It should be an item that can be readily enumerated from the property records and whose specific location can be prescribed from the information shown on the record.
2. To the extent practicable it should be a predominant unit in the sense that it either draws in other classes of property or that it makes up a relatively large part of the investment in the account of which it is a part.

2.4 Sample Size

Distribution of the elements in a class of property with respect to their physical condition is, of course, not known at this stage. Experience indicates, however, that it is conservative to assume a right triangular distribution ranging from 100% condition (new) down to zero percent condition (worn out). The variance (1) of this assumed distribution is,

$$\sigma^2 = \frac{h^2}{18}$$

(h being the range in percent condition) and the precision range of the mean of a random sample of size (n) at the 99.5% assurance level is,

$$r = 2.81 \frac{\sigma}{\sqrt{n}}.$$

Then, the required sample size,

$$n = \frac{(2.81)^2 h^2}{18r^2}.$$

Since h = 100, the respective sample sizes are:

Specified Precision Range at 99.5% Assurance Level (r)	Required Sample Size (n)
$\pm$ 2%	1100
$\pm$ 3%	490
$\pm$ 4%	275

Practical application of these objective sample sizes involves a certain amount of judgment. In Figure 1, it will be noted that the "Aerial Cable" Account is in the upper third of the ranking. Hence about 1100 items should be inspected in this Account. There are 7 distinct property classes in the Aerial Cable Account. The theoretical and actual sample sizes for each of these classes in 3 surveys is indicated in Figure 2.

Figure 2 - SAMPLE SIZES FOR AERIAL CABLE ACCOUNT

Property Class	Decimal Proportion of Total Investment	Theoretical Distribution of a Sample of 1100 Items	Actual Distribution of Sample in 3 Surveys A	B	C
Aerial Cable	.67	739	773	982	720
Aer. Cab. Terms.	.15	165	324	335	236
Strand & Rings	.11	121	768	1014	735
Block Cable	.02	22	95	99	92
Block Cab. Terms.	.01	9	95	99	92
House Cable	.02	22	94	96	90
House Cab. Terms.	.02	22	89	96	90
	1.00	1100	2238	2721	2055

Except for "Aerial Cable" which makes up 2/3rds of the investment in the account, there is evident a strong tendency to over-sample the remaining 6 items. There are several practical reasons for this, viz:

1. The best practical sampling-unit for aerial cable is the "pole-location". Consideration of the Pole Lines Account indicated that a sample of 1500 "Pole Locations" would be required to draw a net sample of 1100 poles (the remainder being locations where attachments were carried on poles of other Companies). On the average, one-half of the "pole locations" carry aerial cable. Hence, a sample of proper size for the Pole Lines Account was also about right for the Aerial Cable Account.
2. The terminals, strand, and rings associated with the aerial cable at a sample pole-location can be inspected at no increase in cost.
3. In the case of block and house cable, the theoretical sample allocation was too small to be meaningful. Hence separate samples of about 100 Blocks and 100 House Cables were drawn so as to obtain a proper cross-section.

The allocation of the sample to Step-by-Step Central Office equipment further illustrates the need for applying judgment to arrive at a practical solution. In one survey 88% of the investment in this account was contained in 7 "large" offices while the remaining 12% of the

investment was scattered over 49 "small" offices. These were sampled separately - 666 items being inspected in the "large" category and 56 in the "small" category. This overall sample resulted in a precision range of ± 3.11% (about half again as wide a range as the objective). The decision to take a sample smaller than the 1100 objective was based on the assumption that the variability between items was likely to be quite small. This assumption proved to be wrong. In fact, had a sample of 1100 been taken, the expected precision range would have been fairly close to the 2% objective, i.e.

$$r' = 3.11 \sqrt{\frac{722}{1100}} = \pm 2.52\%$$

2.5 Selection of Sample Items

The system of sampling found to best meet the rigorous requirements of a property valuation and which, at the same time, is easy to apply may be described as a sub-sampling design - the sample items being selected with a systematic interval from a random starting point. Separation of the total sample into about 10 sub-samples (2), each of which is so selected that it includes items from every section of the area, furnishes a means by which the sample results and their precision range can be computed with a minumum of effort.

A prime requirement for the selection of sample items is the availability of up-to-date records for the respective sampling units. The prescribed number of sampling units are identified from the records by actually counting-off items in accordance with a "Table of Sampling Numbers". A simple example will explain the process.

Suppose a sample of 1500 pole-locations is to be drawn from a total of about 300,000 pole-locations at which a Company owns property. The poles at such locations include fully-owned, partially-owned and foreign Company owned. Assume that the total sample will consist of 10 independent sub-samples of 150 pole-locations each. 300,000 divided by 150 gives a counting interval of 2,000. By choosing 10 numbers at random (from a Table of Random Numbers) in the range between 1 and 2,000 and then applying the counting interval of 2,000 from these starting numbers, a table containing 1500 sampling numbers is provided. A portion of such a table is shown in Figure 3.

Figure 3 - TABLE OF SYSTEMATIC SAMPLING INTERVALS BY SUB-SAMPLES

	SAMPLING NUMBERS					
Sub-Sample	Random Starting Number	2nd	3rd	4th		150th
5	11	2011	4011	6011	...	298,011
3	29	2029	4029	6029	...	298,029
10	231	2231	4231	6231	...	298,231
7	566	2566	4566	6566	...	298,566
1	689	2689	4689	6689	...	298,689
9	841	2841	4841	6841	...	298,841
4	903	2903	4903	6903	...	298,903
8	1201	3201	5201	7201	...	299,201
6	1372	3372	5372	7372	...	299,372
2	1832	3832	5832	7832	...	299,832

The starting numbers are arranged in numerical order to facilitate use of the table. The scrambled order of sub-sample designations results from the fact that they appear in the order that the starting numbers were selected.

To locate the pole locations corresponding to the sampling numbers, the binders containing the pole-line record are assembled and arranged in any order whatever. The pole-locations are then counted-off in sequence, a suitable record being made of the location and sub-sample code as each of the sampling numbers is reached. By using the sampling numbers in sequence by columns, a single counting-off from the record will identify the entire sample. It is also helpful to make a record of certain other information as each sample pole-location is identified, such as: Timber and Treatment; Height and Class of Pole; Ownership; Year Placed; Accounting Classification; Area Code, Etc. Supplemental data of this kind affords a basis for checking the distribution of the sample with the accounting records covering poles.

2.51 Assignment of Sample Items to Inspectors

In a large survey it is preferable that the inspection work be allotted between several inspectors. By assigning the sample items in a definite pattern each sub-sample will include the work of each inspector in approximately equal proportions, and the work of each inspector will in turn consist of an independent sample of property from every section of the area.

For example, if there are 5 inspectors and 10 sub-samples: Inspector A would be assigned to the 1st, 6th, 11th, etc., items in each sub-sample; Inspector B would be assigned to the 2nd, 7th, 12th, etc., items; Inspectors C, D, and E, would likewise be assigned each 5th item.

3.0 INSPECTION PROCEDURE

3.1 Determination of Physical Condition

It is not feasible to measure the percent physical condition of an item of property directly at the time the field inspection is made. The inspection cannot be made under laboratory conditions, in the sense that it would be impractical to provide an inspector with a complete kit of standards by means of which a given item can be compared or gauged. During the inspectors' training course effective use can be made of exhibits of property items mounted on display panels and arranged to show the appearance over the range from "New" to "Worthless". Specifications are then prepared for guidance of the inspectors in the field. These specifications describe the appearance and the number and type of defects which identify each Condition-Grade.

A maximun of five Condition-Grades has been found practicable. It is generally possible to clearly define and identify an item as belonging to some one of five ranges of physical condition. These five Condition-Grades can be generally described as including the following ranges of physical condition:

Condition-Grade	Physical Condition
A	New
B	Good
C	Fair
D	Poor
E	Worn out

Validity of the sample results is mainly dependent upon the uniformity of judgment on the part of the inspectors. Unless it can be demonstrated that a group of inspectors can be depended upon to put items of like condition in the same Condition-Grade the whole process has no value, irrespective of the number of items inspected. While the sample design is laid out in such a way that inspector variability can be identified, yet the success of the survey hinges entirely upon 3 key factors, viz:

1. Use of the smallest practicable number of inspectors, each of whom should be experienced in the installation or maintenance of the property they are to inspect.
2. A week to 10 days of intensive classroom training covering inspection methods, grading of property items and "trial-runs" in which a number of inspectors grade the same set of items.
3. The "multiple" inspection of about 1% of the items in the sample, so arranged that 3 or 4 inspectors grade these particular items.

Limited experience with the reproducibility of results, thru the medium of multiple inspections, shows that 8 out of 10 inspectors will classify a given item in the same grade. The remaining two inspections will split evenly between one grade higher and one grade lower. As the average Physical Condition is unaffected by this distribution the whole process has extremely high validity. In fact, any closer agreement between inspectors would be suspect. Every item inspected occupies some one spot in the range between 100% condition and zero % condition. As this range is sub-divided into a maximum of five parts, some fraction of the property items are more-or-less indeterminate insofar as they belong in one grade or that one adjacent thereto.

3.2 Assignment of Numerical Values to Condition-Grades

For valuation purposes it does not suffice to know how the property is distributed over the Condition-Grades (A, B, C, D, & E). A numerical value (i.e.: percentage) must be assigned to each grade in order to arrive at the property value adjusted for physical deterioration. The determination of sampling precision is likewise based upon percent values of the Condition-Grades. In theory, the sample itself provides the means by which percent values can be assigned to each Condition-Grade. For example, a probalility sample will result in the inspection of wire in every grade in which it exists in the property. These inspections will classify the wire in accordance with the specifications into the various Condition-Grades. Having determined from the property location records the average age of the wire in each Condition-Grade, the scale of percent physical condition is, e.g.

$$\text{Grade A (in percent)} = 100 \left[1- \frac{\text{Average Age of Grade A Wire}}{\text{Average Age of Grade E Wire}} \right]$$

This approach is unduly burdensome. Even where the records do identify

the age of property items the "look-up" process is time-consuming.

This difficulty in the use of a rigorous method for setting percent values for the Condition-Grades has, in general, been overcome by the use of knowledge and experience combined with a liberal amount of judgment. To continue with the case of wire, "Experience" shows the overall physical life to be approximately 20 years. The duration of each Condition-Grade over this overall life was assumed to be:

Condition-Grade	Physical Condition	Age Band	Percent* Condition
A	Bright or obviously new	0 to 3 years	92.5
B	Losing Galvanizing - dull	3 to 11 years	65.0
C	Galvanizing gone (Rust spots)	11 to 15 years	35.0
D	Rusting Generally	15 to 18 years	17.5
E	Pitted (Diameter Reduced)	18 to 20 years	5.0

* Based on mid-point of the age-band.

For certain other classes of property, such as subscribers station apparatus and central office equipment, little knowledge is available about the overall service life. Normal maintenance practice tends to hold such property in a high state of serviceability through piece-part replacements. In this case it can be assumed that some lower limit (say 40%) is the lowest physical condition that will be encountered for any item yet in service. The range between 100% and 40% is then arbitrarily apportioned to whatever number of Conditioned-Grades are deemed appropriate.

3.21 Some Further Observations on Assigning Numerical Values to Condition-Grades

The net outcome of these empirical approaches to numerical evaluation of Condition-Grades is a rather extensive array of values, as between property classes, for any given Condition-Grade. The values are, of course, no better than the judgment that went into their formation. In fact the "experts" show a disconcerting variation in judgment on this score. Not only do they tend to differ rather widely in opinions regarding overall life, but display a fairly wide range of judgment as to the proportion of overall life encompassed by a given Condition-Grade. Here, then, is a paradox. The Condition-Grade, as determined by field inspection, is highly reliable. On the contrary, the percent value of a given Condition-Grade, - as determined by the "experts" - tends to be heterogeneous.

There is room for possible controversy on this matter of estimating overall service life. An approach that avoids this area of uncertainty is suggested in the examples that follow.

In view of the fact that the whole process involved in the measurement of physical condition is a subjective one - in the sense that judgment is required - the first place to eliminate variability is in the specifications for the Condition-Grades. A convenient way to accomplish this is to define the middle of the range (i.e. grade C) in such a manner that it corresponds to a numerical condition of 50%. As grades B and D are intuitively half-way between grade C and the ends of the scale of condition. no special difficulty is presented in specify-

ing the physical condition for these intermediate points, once a satisfactory grade C specification is arrived at. In short, the definitions of the Condition-Grades should convey the same concept of percent physical condition for each class of property.

Having taken this step, the way is then clear to develop a uniform set of percent values that apply to each property class. This can be done by continuing the subjective process. Let the range of defects that correspond to Condition-Grade A be (0 to d) defects; for Condition -Grade B, (d to 2d) defects - and so on. If the mid-points of these defect ranges are plotted along the X-axis and ordinates erected thereat whose height is

$$Y = e^{-\frac{X^2}{c}},$$

a smooth curve can be constructed which will indicate the physical condition (Y) for any number of defects (X). The slope of a "die-away" curve of this type is such that it reflects a lower rate of deterioration during the first half of the life of an item of property than during the remainder of its life.

The development of such a Condition Grade - Percent Condition curve first requires solution for the parameter (c) using known values for (X) and (Y). In this case, Y = .5 (50% condition) when X = 2.5d (mid-point of the range between 2d and 3d defects).

Expressing the equation in logarithmic form,

$$\log Y = -\frac{x^2}{c} \log e$$

$$\text{and } c = -x^2 \frac{\log e}{\log Y}$$

$$\text{Now, } x^2 = (2.5d)^2 = 6.25d^2$$

$$\log e = .4343$$

$$\log Y = (-.301)$$

Substituting,

$$c = 6.25d^2 \frac{(.4343)}{(.301)}$$

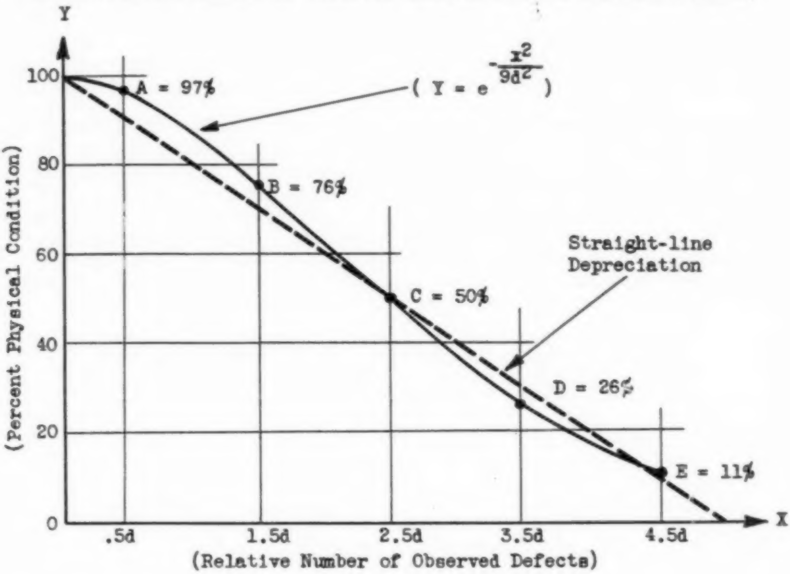
$$c = 9d^2$$

Using $c = 9d^2$, the value of Y is then computed for each Condition-Grade, as shown in Figure 4.

It can be argued, of course, that the shape of this curve does not exactly conform to that which might be expected for some particular class of plant. Creosoted pine poles, for example, can be taken to be Grade A, irrespective of their age, as long as there is no evidence that the preservative has disappeared. When the creosote is gone, however, the pine timber is subject to very rapid decay in most localities. Hence some middle-grade may be more theoretical than actual, as the timber (based on appearance of preservative) may be definable in Condition B and Condition D, but not in Condition C. In any such case the way to get over the hurdle is simply to jump it. If it is not possible to define the defects for certain intermediate Condition-Grades, then do not try - just leave them out.

Figure 4 - CONDITION-GRADE - PERCENT CONDITION CURVE

(Same relative number of observed defects in each Condition-Grade)



Condition-Grade	No.of Defects (x)	x^2	$(- \frac{x^2}{c})$		log y	y
A	.5d	$.25d^2$	- .0278	Mul.by log e=	-.0121	.97
B	1.5d	$2.25d^2$	- .2500	" " " "	-.1084	.76
C	2.5d	$6.25d^2$	- .6950	" " " "	-.3015	.50
D	3.5d	$12.25d^2$	-1.3615	" " " "	-.5910	.26
E	4.5d	$20.25d^2$	-2.2500	" " " "	-.9760	.11

It is not possible to claim any greater accuracy for this simplified approach to the problem of setting percent values to the Condition-Grades. The values suggested, however, conform fairly closely to the median values used in five recent telephone rate cases. Hence the average physical condition of the property seems quite likely to be closely identical to that obtained by a more round-about approach. The main advantage in the use of a uniform scale is that it conforms to percent values by Condition-Grades that would be expected when property is classified between 5 levels of physical condition. It completely avoids any possible field of controversy between opposing experts as to the overall life of property items.

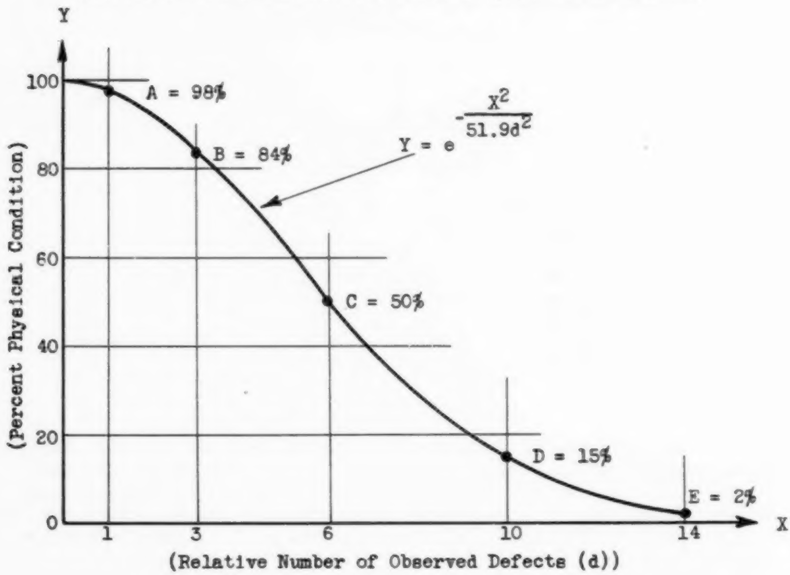
There is also a distinct advantage toward assuring uniformity of judgment during the field inspection work. When the Condition-Grade specifications vary widely between property classes with respect to the severity of defects an inspector is quite likely to develop a bias. For example, if the observable defects in one class of property are so

defined that Grade C represents a 20% physical condition, but for some other property class Grade C corresponds to a 60% physical condition, the inspector has a natural inclination to interpret his instructions in such a way that Grade C is <u>intended</u> to mean the same degree of deterioration in both cases. This can result in either an upward or downward bias of unknown degree. The use of a uniform scale of Condition-Grades avoids the possibility of serious inspector bias.

It should be pointed out that the "slope" of this curve can be adjusted to meet any specified scale of defects. Expert judgment may, for example, bring to light that the curve should drop less sharply in the A and B range, but more sharply in the D and E range. All that is required to develop a new curve is to determine what scale of relative defects should apply to each Condition-Grade. Suppose the two top grades are allowed the same number of defects but the three lower grades are each judged to show progressive deterioration at about twice this rate. Figure 5 illustrates this situation.

Figure 5 - CONDITION-GRADE - PERCENT CONDITION CURVE

(Variable Number of Observed Defects by Condition-Grades)



Another variation is to use any one of the Condition-Grades as a bench-mark. For example, if Grade E is set at 10% condition with the same relative scale of defects shown in Figure 5, the values of the Condition-Grades would be defined as: A = 99%; B = 90%; C = 66%; D = 31%; E = 10%.

4.0 COMPILATION OF RESULTS, INCLUDING PRECISION OF MEASUREMENT

A well-designed process for tabulating, summarizing and processing the field inspectors' results is an important step in the survey. The methods outlined herein have been used with success with manual tabulating. Where punch card equipment is available the same general prodedure applies as the operations can be readily coded for punch card use.

A Summary form that shows for each class of property the distribution of sample items by Condition-Grades, Inspectors, and Sub-Samples is the basic requirement. It is unnecessary to provide a form for preliminary summarization as the inspection forms themselves can be readily sorted, 1st by Sub-Sample, 2nd by Inspector, and 3rd by Condition-Grade. When a given field inspection form covers more than one property class, the sorting technique is, of course, applied in sequence for each such class. A summary form such as Figure 6 has been found satisfactory.

Figure 6 - FORM FOR INITIAL SUMMARY OF SAMPLE DATA

ACCOUNT_____ SUB-SAMPLE_________TYPE OF PLANT_______________________

DECIMAL PROPORTION OF TOTAL INVESTMENT IN ACCOUNT____________________

Condition Grade	Percent Condition	Distribution of Sample Items by Inspectors					Sums & Averages
		I	II	III	IV	V	
A							
B							
C							
D							
E							
Sums							
Percent Aggregates							
Average Physical Condition							

The average Physical Condition (computed on the last line of each Sub-Sample Section) is then entered on a computing form for the determination of the overall physical condition, its range of precision and analysis of variance.

Figure 7 illustrates the method used for determining the average physical condition and the sampling precision for a given class of property. Sample data for the entire property (to which sampling inspection was applied) is shown in Figure 8.

References: (1) Some Theory of Sampling - Deming - Page 52.
(2) Some Theory of Sampling - Deming - Page 352.

INDICATED PRECISION OF SAMPLING

POLE LINES (ACCT. 241)

Line	(a) Item	(b) Average % Condition By Sub-Samples (X_1) 1	2	3	4	5	6	7	8	9	10	(c) Decimal Proportion of Total Account (W_1)	(d) No. of Items in Sample (N_1)
1	Poles - Fully Treated	74.4	73.6	74.7	74.3	75.3	74.6	77.3	76.2	76.1	75.2	.5558	992
2	Poles - Partially Treated	36.9	33.4	38.3	32.4	43.0	34.5	40.5	37.7	44.6	41.9	.0583	116
3	Poles - Untreated	5.0	10.0	10.0	47.0	51.5	34.0	10.0	34.5	20.0	5.0	.0290	26
4	Anchor Rods	78.5	76.3	76.2	76.9	80.2	74.7	75.7	75.5	73.8	74.7	.0817	308
5	Guys	73.9	68.5	71.7	70.9	77.3	67.1	71.6	74.3	68.0	72.3	.1304	513
6	Cross-arms	53.8	52.5	51.8	54.1	57.1	63.1	58.2	55.1	59.8	56.1	.1448	949
7	Avg. of Sub-Sample Means = $\sum_1^6 X_1 W_1$	67.5	65.9	67.1	67.9	70.8	68.4	69.6	69.4	69.0	68.0	1.0000	2904
8	Average of Sub Sample Means = $\overline{X} = (\sum_1^{10}\sum_1^{6} X_1W_1) \div 10 = 68.4\%$												
9	Estimated Variance of the Mean = $\sigma_{\overline{X}}^2 = \sum_1^{10} (\sum_1^{6} X_1W_1 - \overline{X})^2 \div 90 = .159$												
10	Estimated Standard Error of the Mean = $\sigma_{\overline{X}} = \sqrt{.159} = .40\%$												
11	Precision Range of Average Physical Condition = $r = 3\sigma_{\overline{X}} = \pm 1.2\%$												

Figure 7 - PERCENT PHYSICAL CONDITION FOR POLE LINE ACCOUNT

INDICATED PRECISION OF SAMPLING

SUMMARY

		a	b	c	d	e	f	g
Line	Property Account	Description	Number of Items Inspected (N)	Average Per Cent Condition ($\bar{X}$)	Decimal Proportion of Total Plant Sampled (W)	Estimated Standard Error of Average % Condition ($\sigma\bar{X}$)	Weighted Variance of Average % Condition $(W\sigma\bar{X})^2$	Indicated Precision Range of Average % Condition ($\pm$3xCol.E)
1	221	C.O.E. Manual	1,192	93.0	.0299	.22%	.000043	.66%
2	221	C.O.E. Panel	2,482	94.0	.0988	.20	.000390	.60
3	221	C.O.E. Cross-bar	985	96.4	.1117	0	0	0
4	221	C.O.E. Step-by-Step	595	92.6	.0143	.09	.000002	.27
5	221	C.O.E. Toll-Except Switchboard	29	93.6	.0130	.65	.000071	1.95
6	231	Station Apparatus	1,577	85.5	.0518	.30	.000241	.90
7	232	Station Installations	1,359	86.2	.0520	.60	.000973	1.80
8	233	Drop & Block Wires	3,225	74.3	.0302	.49	.000219	1.47
9	234	Private Branch Exchange	705	87.9	.0413	.87	.001291	2.61
10	235	Booth & Special Fittings	56	83.0	.0075	1.28	.000092	3.84
11	241	Pole Lines	2,904	68.4	.1045	.40	.001747	1.20
12	242.1	Aerial Cable	2,285	81.9	.1726	.29	.002505	.87
13	242.2,3,4	Underground, Buried & Submarine Cable	443	79.8	.1177	.50	.003463	1.50
14	243	Aerial Wire	8,123	66.4	.0305	.61	.000346	1.83
15	244	Underground Conduit	2,668	93.0	.1042	.20	.000434	.60
16	264	Vehicles & Other Work Equipment	1,580	79.8	.0200	1.26	.000635	3.78
17		Total Plant Conditioned By Sampling	30,208	84.8	1.0000	.11%*	.012452	$\pm$.33%

* Column E, Line 17 = Square Root of Column F, Line 17.

Figure 8 - SUMMARY OF PERCENT PHYSICAL CONDITION BY ACCOUNTS

EXPERIMENTS COMPARING TWO METHODS FOR PERCENTAGE DEFECTIVE

Frank Proschan
Sylvania Electric Products, Inc.

Abstract

This paper describes a convenient statistical test of significance to help evaluate the outcome of experiments in which two methods, materials, products, or, in fact, groups of any kind are being compared on the basis of percentage defective or, in general, percentage possessing any specified attribute. In addition, a nomograph is presented to enable the experimenter to estimate the number of observations required to reach conclusions with any specified degree of precision from such experiments.

Introduction

A very common type of experiment performed in the development of a new product or the improvement of a standard product involves observing the percent defective (or possessing any specified attribute) for the standard product and the corresponding percent defective for the experimental product. If the two percentages differ "enough", the conclusion is reached that the experimental variation has affected product quality. If the difference is not large enough, the inference is made that no conclusive evidence exists that the standard and the experimental products really differ in long-run percentage defective.

But what must the minimum difference in observed percentages be, to permit a reliable conclusion that the two product variations really differ in quality? Clearly, the minimum difference required for significance depends on the number of units tested and the percentage defective observed. For example, an observed difference of 10% might be conclusive for 1000 units of each product type, but quite insignificant for only 10 units of each product type. Also, a difference of 10% might be conclusive if 10% defective had been observed for the standard product and 0% defective observed for the experimental type, while the same 10% difference might be unconvincing if the actual percentages observed were 55% vs. 45%.

It is apparent, then, that an objective statistical test is required to evaluate the significance of the observed difference in percentages. In this paper the required statistical test of significance, the chi-square test, is explained and illustrated.

I am indebted to Professor Sir Ronald A. Fisher, Cambridge, to Dr. Frank Yates, Rothamstad, and to Messrs. Oliver and Boyd Ltd. Edinburgh, for permission to reprint Table No. IV from their book "Statistical Methods for Biological, Agricultural, and Medical Research", and to Professors Paulson and Wallis and the McGraw-Hill Book Co. for permission to adapt the nomograph of Chap. 7, "Techniques of Statistical Analysis" edited by Eisenhart, Hastay, and Wallis.

Problem

Consider the following typical problem. Plant X is developing a new product by hot-drawing methods. An experiment is performed to determine whether two die lubricants differ in their effect on surface quality. The results obtained are:

Table 1. Observed Frequencies

	Lubricant 1	Lubricant 2	Total
Number of defective items	54	47	101
Number of acceptable items	105	114	219
Total	159	161	320

Thus, the use of lubricant 1 resulted in a percent defective of 34.0%, while the use of lubricant 2 resulted in a percent defective of 29.2%. Does this observed difference of 4.8% stem from a real difference in lubricant effect on quality, or it is simply the result of sampling or chance fluctuations?

Solution

Suppose the two lubricants were really alike in percent defective in the long run. What magnitude of difference in the two observed percentages defective would be readily explainable on the basis of chance fluctuations?

To answer this question, proceed as follows. If the two lubricants were really alike in long run percentage defective, the data for both lubricants would be pooled together to estimate the common defective rate as $\frac{101}{320} = 0.316$. Therefore, for the 159 lubricant 1 items, an estimated $0.316 \times 159 = 50.2$ defectives would result. Similarly, a table of frequencies expected on the hypothesis that both lubricants are alike in long-run percent defective could be constructed, as follows:

Table 2. Expected Frequencies

	Lubricant 1	Lubricant 2	Total
Number of defective items	50.2	50.8	101
Number of acceptable items	108.8	110.2	219
Total	159.0	161.0	320

The remaining values in the table, 50.8, 108.8 and 110.2 are most readily obtained by difference; e.g., 108.8 for the number of good items obtained using lubricant 1 is derived as the difference between the total of 159 lubricant 1 items and the estimated 50.2 defective lubricant 1 items.

The original question: "Do the two lubricants really cause a difference in percent defective?" now becomes, "Are the observed values (Table 1) significantly deviant from the corresponding expected values (Table 2)?" In this form, the question may be readily answered by the application of a simple statistical test known as the "chi-square" test, which, in general, helps determine whether a set of observed frequencies departs significantly from the corresponding expected or theoretical frequencies.

To apply the test, calculate χ^2 (chi-square) from

$$\chi^2 = \Sigma \frac{\{|\sigma - E| - 0.5\}^2}{E} \qquad (1)$$

where

σ = observed value
E = expected value
Σ represents the sum to be obtained for the four entries of Tables 1 and 2.

Substituting,

$$\chi^2 = \frac{\{|54 - 50.2| - 0.5\}^2}{50.2} + \frac{\{|47 - 50.8| - 0.5\}^2}{50.8} + \frac{\{|105 - 108.8| - 0.5\}^2}{108.8} +$$

$$\frac{\{|114 - 110.2| - 0.5\}^2}{110.2} = 3.3^2\ (1/50.2 + 1/50.8 + 1/108.8 + 1/110.2)$$

$$= 0.631$$

Next compare the computed value of χ^2, 0.631, with the entries of Table A to determine the likelihood of obtaining a value of χ^2 as large as 0.631 on the assumption of no difference between lubricants. Note that 0.631 falls between the tabular values 0.455 and 1.07; hence the probability is between 0.50 and 0.30 of obtaining by chance a value of χ^2 as large as 0.631 on the assumption of no difference between lubricants. Therefore the observed outcome of the experiment does not render untenable the assumption of no difference between lubricants, since the value of χ^2 actually obtained under this assumption was not at all unlikely. In other words, no conclusive evidence exists that the two lubricants really differ in their long run proportion defective.

Rationale

The logic underlying the solution may be outlined thus:

Question: Are the two lubricants different in long run proportion defective?

Solution: Assume the two lubricants have the same proportion defective in the long run.

1. Calculate χ^2 on this assumption.
2. From Table A determine the probability of obtaining as large a value of χ^2 by chance.
3. If the probability is small, conclude that the assumption is untenable; i.e., conclude that the two lubricants differ in long run proportion defective. Otherwise, infer that no conclusive evidence exists that the two lubricants differ.

Risks

How small must the probability be for the assumption of no difference between the two lubricants to be rejected? Conventionally, a probability level (in statistical jargon, "significance level") of 0.05 is used; i.e., the 0.05 value of χ^2 in Table A, 3.84, is taken as the critical value - a computed value as large as 3.84 implying a difference between the two lubricants; a computed value smaller than 3.84 implying a lack of conclusive evidence of such a difference.

Of course, the use of the 0.05 level of significance is arbitrary. If, in a particular problem, it is important to keep to an absolute minimum the risk of falsely concluding that the two materials or methods differ, then the 0.010 or even the 0.005 level may be used. On the other hand, in some situations it will be sufficient to set the level of significance at 0.10 or 0.20.

Suppose now the 0.05 point of χ^2, 3.84, is being used as the critical value. As stated above, this means that if the two materials are alike in proportion defective, only 5 times in 100 will the value of χ^2

calculated from the experimental data be significantly large, i.e., as large as 3.84. But on those 5 occasions, the significantly large value of χ^2 will lead to the conclusion (falsely) that the two materials differ, when actually they are alike in proportion defective. In other words, the risk of falsely concluding that the two materials differ when they do not is 5 in 100 if the .05 level of significance is used.

What about the risk of falsely concluding that the two materials are alike in long run percent defective when actually they differ? Clearly, this second type of risk will depend on the amount of difference actually existent between the two percentages defective - if the two materials have very different percentages defective, the risk of a false conclusion is very small; if the two materials have percentages defective only slightly different, the risk of a false conclusion is quite large (close to 95 in 100 if the 0.05 level of significance is being used).

Special Form For χ^2- Computations

Because of the large number of comparisons being made at the Hicksville Pilot Plant as to the percentage defective of two methods or materials, it was found quite helpful to prepare a form outlining the computational steps required to reach a conclusion using the χ^2-test. The form is displayed below. The form has been used successfully by a number of engineers after a preliminary explanation of the rationale and the technique of the χ^2-test.

- -

Comparison of Two Percentages Using the χ^2-Test

Observed Frequencies

	Group I	Group II	Total
Good	(1)	(3)	(5)=(1)+(3)
Bad	(2)	(4)	(6)=(2)+(4)
Total	(7)=(1)+(2)	(8)=(3)+(4)	(9)=(5)+(6) =(7)+(8)

Theoretical Frequencies (1 decimal place)

	Group I	Group II	Total
Good	(10)=(5)(7)/(9)	(12)=(5)-(10)	(5)
Bad	(11)=(7)-(10)	(13)=(8)-(12) =(6)-(11)	(6)
Total	(7)	(8)	(9)

$$\chi^2 = \left\{\left|(1) - (10)\right| - 0.5\right\}^2 \left\{\frac{1}{(10)} + \frac{1}{(11)} + \frac{1}{(12)} + \frac{1}{(13)}\right\}$$

$\chi^2 =$

where (1) - (10) represents the absolute value of the difference between (1) and (10).

CROSS OUT ONE:

$\chi^2 \geq 3.841$; therefore Groups I and II are different in % defective.
$\chi^2 < 3.841$; therefore no conclusive evidence exists that Groups I and II are different in % defective.

NOTE:

1. To use this technique, (10), (11), (12), and (13) must be ≥ 5.
2. The critical value of χ^2, 3.841, is that value which will be exceeded by chance 1 time in 20 when the 2 groups are actually alike in % defective.
3. For more information on the χ^2-test, see Eisenhart, Hastay, and Wallis, "Techniques of Statistical Analysis", McGraw-Hill Book Company, 1947, Chapter 7.

Issued by: Quality Control Department
Sylvania Electric Products
Hicksville, New York

- -

Formula for χ^2

In actual practice, some users may find it more convenient to use a formula for χ^2 especially suited for the present problem of comparing two percentages. Let the observed frequencies be symbolized as in the following table:

Table 3. Observed Frequencies

	Material 1	Material 2	Total
Number of defective items	a	b	a + b
Number of acceptable items	c	d	c + d
Total	a + c	b + d	a + b + c + d = T

Then χ^2 may be computed from

$$\chi^2 = \frac{T\left\{|ad - bc| - T/2\right\}^2}{(a+b)(a+c)(b+d)(c+d)} \qquad (2)$$

Note that the factors in the denominator are the marginal totals, while T is the grand total.

It can easily be shown that formulas (1) and (2) are algebraically equivalent. Thus, in the above illustration using formula (2) leads to

$$\chi^2 = \frac{320\left\{|54 \times 114 - 47 \times 105| - 320/2\right\}^2}{101 \times 159 \times 161 \times 219} = 0.631,$$

which is the same as the value obtained by using formula (1).

Experiments To Detect Superiority

The experiment described above was performed essentially to determine whether a difference actually existed in percentage defective for the two lubricants. In many situations, however, experiments will be performed

to determine whether one material or method is superior to another. Questions of superiority would arise, for example, if one were comparing a new type of lubricant with the standard type; in such a case, one is really interested in determining whether the new type is superior, not just whether it is different.

For experiments to detect a superiority, the χ^2-test may be used just as before. Only now the first type of risk, that of falsely concluding that the new lubricant is superior, is half the previously discussed risk of falsely concluding that the lubricants are different. This follows from the fact that in the previous case of testing for a difference, the two lubricants could appear different in two ways: lubricant 1 might have appeared superior or lubricant 1 might have appeared inferior. The second type of risk, that of falsely concluding that the two lubricants are alike, changes only slightly however.

Comments

1. To use the χ^2-test of significance as explained above, it must be assumed that the probability of producing a defective by either of the two methods being compared is constant throughout the experiment. In other words, production of units of both types must be controlled during the experiment.

2. When any of the four theoretical or expected values turns out to be smaller than 5, the χ^2-test described above is no longer appropriate. Other procedures are discussed in reference 2.

3. Compute the four theoretical or expected values to one decimal place. Additional accuracy is not usually required.

4. Suppose now a comparison is to be made among more than two groups with regard to percentage defective. Is the χ^2- test applicable? Yes! The details are presented in reference 4, pp. 204-209. In fact, the χ^2-test may be used for a two-way classification with any number of categories in either direction.

5. An unusually small value of χ^2 indicates excessive agreement between the observed and the corresponding theoretical frequencies. For example, in the lubricant comparison experiment described above, suppose the calculated value of χ^2 had turned out to be 0.003. Then Table A shows that the probability of obtaining as small a value of χ^2 as 0.003 is between 2 in 100 and 5 in 100. In other words, the agreement between the lubricant 1 and lubricant 2 failure rates is so good as to occur only between 2 and 5 times in 100 even when the two lubricants are alike in percentage defective. Such unusually good agreement might have arisen from one or more of the following causes: a) doctored data b) biased measurements c) the operation of extraneous factors.

Size of Sample Needed

Suppose, now, that one is planning an experiment to determine whether a new method or material is superior to the standard method or material. How many observations should one plan to make on each of the two methods or materials to determine if they really differ in percentage defective?

The answer depends on the risks one is willing to take and on the

actual difference between the unknown percentages defective. In a general way, the smaller the risks permitted, the larger is the number of observations needed; also the smaller the difference between the unknown percentages defective, the larger must be the number of observations needed to detect this difference. To get a specific answer, one may use the nomograph below to estimate N, the required sample size using each method, as follows.

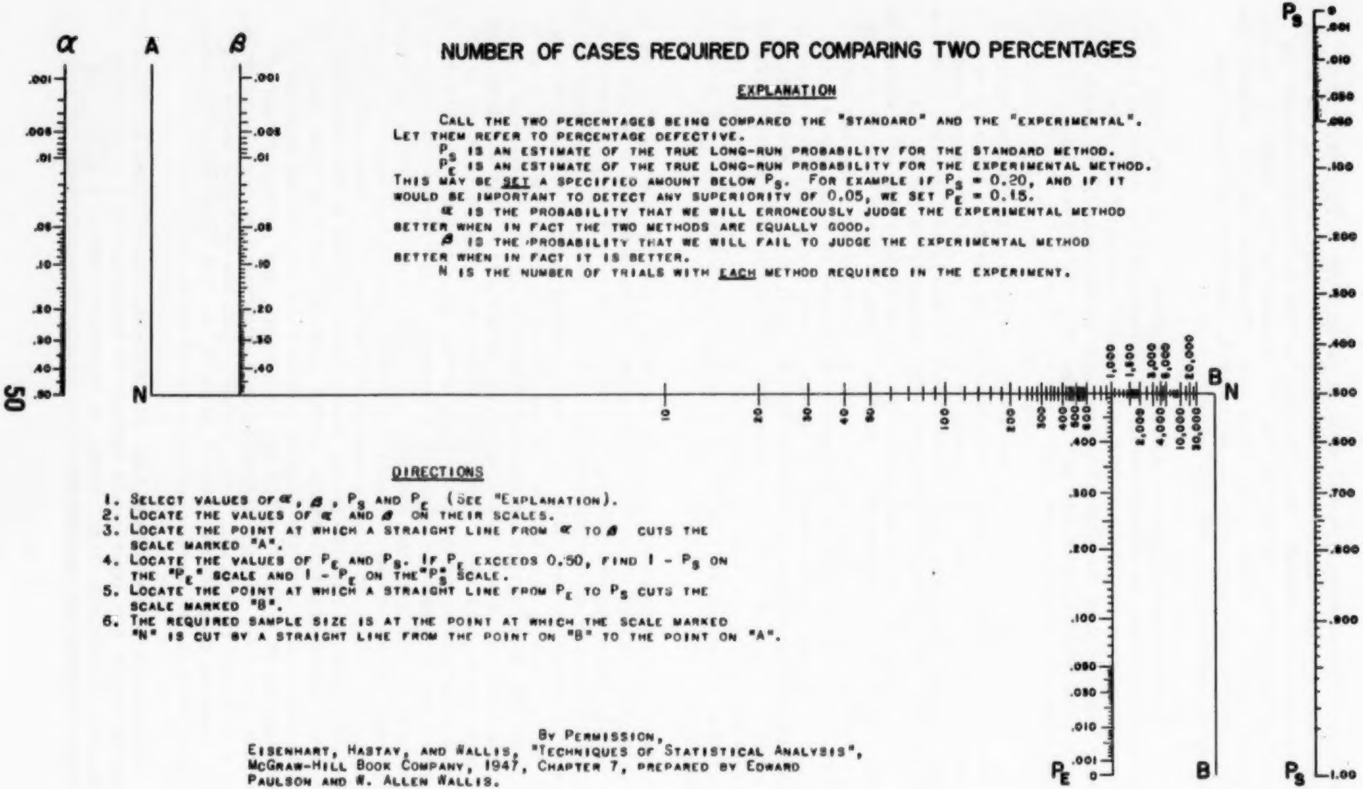
For example, suppose an electronic laboratory is developing a new model of a tube. An experiment is to be performed to determine whether the new version is superior to the standard version of the tube. How many tubes of each type shall be compared experimentally to reach a reliable conclusion as to whether the new tube has a lower long-run failure rate than the standard tube?

Suppose past experience has shown a failure rate for the standard tube, P_S, of 0.20. Suppose, too, that if the new tube has a 5% or greater superiority over the standard tube (i.e., if the experimental tube failure rate, P_E, is 0.15 or less), it would be important that the experiment reveal this superiority. Therefore P_E is set at 0.15. Finally, suppose the laboratory wishes to tolerate risks of a wrong conclusion of at most 10 in 100. In other words, α, the risk of falsely concluding that the experimental tube is superior when the two tube models are really alike in proportion defective is 0.10; and β, the risk of falsely concluding that the two tubes are alike, when actually the experimental tube is superior by 5%, is also 0.10. $\alpha = 0.10$ is connected with $\beta = 0.10$ to determine a point on A; also $P_S = 0.20$ is connected with $P_E = 0.15$ to determine a point on B. The A and B points are then connected to locate the required sample size for <u>each</u> tube model, N = 600.

For some experiments, it will be found that for specified values of P_S, P_E, α, and β, the required number of observations N determined from the nomograph will be larger than can be practically obtained because of time or money limitations. In such cases, the nomograph may be used to determine the risks of using a smaller number of observations. As before, the P_S and P_E values are joined to determine a point on B. A straight line is then extended through the point on B and the number of observations N that are <u>practically</u> attainable. This straight line intersects a point on A. From this point on A, the risks α and β of false conclusions may be determined as any pair of values on the α and β scales that may be joined by a straight line through the point on A. For example, if 600, the required number of tubes of both the standard and experimental types in the illustration just above, were too large to test practically, it might be realistic to specify a practical number N = 200 tubes to be made by the standard method and by the experimental method. Recalling that values of $P_S = 0.20$ and $P_E = 0.15$ were used, the risks associated with the smaller sample size N = 200 may then be determined as $\alpha = 20\%$, $\beta = 27\%$ or, in general, any pair of α and β values that pass through the point on A determined by the values N=200, $P_S = 0.20$ and $P_E = 0.15$. Note, though, that the <u>critical</u> value of χ^2 that must be used for α to be 0.20 is $\chi^2 = 1.64$ (see Table A) rather than $\chi^2 = 3.84$, which was required for α to be 0.05.

It should be emphasized that the nomograph is designed to yield a sample size N for experiments seeking to determine if a <u>superiority</u> exists. If the experiment seeks to determine whether a <u>difference</u> exists between the two methods or materials compared, then the only

NUMBER OF CASES REQUIRED FOR COMPARING TWO PERCENTAGES



EXPLANATION

Call the two percentages being compared the "standard" and the "experimental". Let them refer to percentage defective.

P_S is an estimate of the true long-run probability for the standard method.

P_E is an estimate of the true long-run probability for the experimental method. This may be <u>set</u> a specified amount below P_S. For example if $P_S = 0.20$, and if it would be important to detect any superiority of 0.05, we set $P_E = 0.15$.

α is the probability that we will erroneously judge the experimental method better when in fact the two methods are equally good.

β is the probability that we will fail to judge the experimental method better when in fact it is better.

N is the number of trials with <u>each</u> method required in the experiment.

DIRECTIONS

1. Select values of α, β, P_S and P_E (See "Explanation).
2. Locate the values of α and β on their scales.
3. Locate the point at which a straight line from α to β cuts the scale marked "A".
4. Locate the values of P_E and P_S. If P_E exceeds 0.50, find $1 - P_S$ on the "P_E" scale and $1 - P_E$ on the "P_S" scale.
5. Locate the point at which a straight line from P_E to P_S cuts the scale marked "B".
6. The required sample size is at the point at which the scale marked "N" is cut by a straight line from the point on "B" to the point on "A".

By Permission,
Eisenhart, Hastay, and Wallis, "Techniques of Statistical Analysis", McGraw-Hill Book Company, 1947, Chapter 7, prepared by Edward Paulson and W. Allen Wallis.

change is that the nomograph is entered with 1/2 the desired value of α.

Pooling Data

Statement 1 of "Comments" above specified that the χ^2-test as described was strictly applicable only to situations in which the probability of producing a defective by either of the two methods being compared remained constant throughout the experiment. Unfortunately, in many situations this may not be true. For example, in the lubricant experiment described above, suppose half of the units produced (using each lubricant) came from one lot of raw material, and the other half came from a different lot of raw material. Now if raw material were an important variable affecting surface quality (the quality characteristic under study), it is evident that the probability of producing a defective would not remain constant for each lubricant throughout the experiment. Can the χ^2-test be modified to be applicable to this situation?

Yes, by a simple device. Separate the data into two sets, one representing the first lot of raw material, and the other representing the second lot of raw material, as follows:

Table 4. Observed Frequencies

First Batch of Raw Material

	Lubricant 1	Lubricant 2	Total
Number of defective items	25	22	47
Number of acceptable items	55	59	114
Total	80	81	161

Second Batch of Raw Material

	Lubricant 1	Lubricant 2	Total
Number of defective items	29	25	54
Number of acceptable items	50	55	105
Total	79	80	159

Then proceed as before for each set of data separately:

Table 5. Expected Frequencies

First Batch of Raw Material

	Lubricant 1	Lubricant 2	Total
Number of defective items	23.4	23.6	47.0
Number of acceptable items	56.6	57.4	114.0
Total	80.0	81.0	161.0

Second Batch of Raw Material

	Lubricant 1	Lubricant 2	Total
Number of defective items	26.8	27.2	54.0
Number of acceptable items	52.2	52.8	105.0
Total	79.0	80.0	159.0

Next calculate χ^2 for each set of data using formula (3), which omits the 0.5 correction factor of formula (1):

$$\chi^2 = \Sigma \frac{(O - E)^2}{E} \qquad (3)$$

where, as before, O represents the observed frequency, and E the corresponding expected frequency on the hypothesis of no difference between lubricants.

For the first raw material lot,

$$\chi^2 = (25 - 23.4)^2 \left(\frac{1}{23.4} + \frac{1}{23.6} + \frac{1}{56.6} + \frac{1}{57.4}\right) = 0.308 \;;$$

for the second lot

$$\chi^2 = (29 - 26.8)^2 \left(\frac{1}{26.8} + \frac{1}{27.2} + \frac{1}{52.2} + \frac{1}{52.8}\right) = 0.543$$

At this stage, a convenient property of χ^2 is utilized. The χ^2 statistic is additive [4,p.189]; that is, the computed values of χ^2 for the two lots of raw material may be added together, giving

$$\chi^2_{\text{Combined}} = 0.308 + 0.543 = 0.851$$

The χ^2_{Combined} value is, therefore, a measure of the departure of the observed frequencies from the corresponding theoretical frequencies for <u>both</u> sets of data. However, it would be incorrect to use Table A to evaluate the significance of a computed value of $\chi^2_{\text{Combined}} = 0.851$, since Table A contains critical values of χ^2 calculated from <u>one</u> double dichotomy, not for the sum of several such χ^2 values.

Table B must therefore be used. For the present problem, the 0.05 level of significance of χ^2_{Combined} calculated as the total of the χ^2's for the <u>two</u> double dichotomy tables is 5.99. Since the calculated value of χ^2 is <u>only</u> 0.851, which is less than the critical value of χ^2, 5.99, no conclusive evidence exists that the two lubricants are different in their reject rate.

In general, then, if the data comparing the two methods originally are obtained under two or more sets of conditions, it is necessary to compute a separate χ^2 value from formula (3) for each set of circumstances, and then to add the χ^2 values together. The total χ^2 value is then compared with the appropriate critical value of Table B to determine if the two methods really differ or not.

Conclusion

The experimenter need no longer scratch his head, sigh, and wonder what the percentages observed for the two methods, processes, materials, etc., really prove. By using the χ^2 test properly, he can objectively and quantitatively determine the significance of the difference in observed percentages, and reach rational decisions accordingly.

The test may be applied to a wide variety of problems in diversified fields. Thus, comparisons of 2 raw materials, 2 machines, 2 methods of production, 2 operators, 2 companies, 2 lots, etc., may be performed in

the chemical, mechanical, aircraft, pharmaceutical, food-processing, etc., industries. The important restriction is that the data must be in the form of frequencies rather than measurements.

REFERENCES

1. Dixon and Massey, "Introduction to Statistical Analysis", McGraw-Hill Book Co., 1951, pp. 188-190.

2. Eisenhart, Hastay, and Wallis, "Techniques of Statistical Analysis", McGraw-Hill Book Co., 1947, pp. 249-265.

3. Li, C. H., "A Nomograph for Evaluating the Significance of Some Test Results", ASTM Bulletin, December 1953, pp. 74-76.

4. Snedecor, G. W., "Statistical Methods", The Iowa State College Press, 1946, pp. 194-200.

Table A. Percentage Points of the χ^2 Distribution For Comparing Two Percentages

α, Probability of exceeding critical value of χ^2	Critical Value of χ^2
0.99	0.0002
0.98	0.0006
0.95	0.004
0.90	0.016
0.80	0.064
0.70	0.148
0.50	0.455
0.30	1.07
0.20	1.64
0.10	2.71
0.05	3.84
0.02	5.41
0.010	6.63
0.001	10.8

For example, the probability is 0.50 of obtaining a calculated value of χ^2 as large as 0.455 in sampling from two populations having the same percentage defective.

Table A is abridged from Table IV of Fisher and Yates: Statistical Tables for Biological, Agricultural, and Medical Research, published by Oliver and Boyd Ltd. Edinburgh, by permission of the authors and publishers.

Table B. Values of $\chi^2_{combined}$

Number of χ^2 values combined	0.50	0.30	0.20	0.10	0.05	0.02	0.01
	Probability of a Larger Value of $\chi^2_{combined}$						
1	0.46	1.07	1.64	2.71	3.84	5.41	6.64
2	1.39	2.41	3.22	4.60	5.99	7.82	9.21
3	2.37	3.66	4.64	6.25	7.82	9.84	11.34
4	3.36	4.88	5.99	7.78	9.49	11.67	13.28
5	4.35	6.06	7.29	9.24	11.07	13.39	15.09
6	5.35	7.23	8.56	10.64	12.59	15.03	16.81
7	6.35	8.38	9.80	12.02	14.07	16.62	18.48
8	7.34	9.52	11.03	13.36	15.51	18.17	20.09
9	8.34	10.66	12.24	14.68	16.92	19.68	21.67
10	9.34	11.78	13.44	15.99	18.31	21.16	23.21
11	10.34	12.90	14.63	17.28	19.68	22.62	24.72
12	11.34	14.01	15.81	18.55	21.03	24.05	26.22
13	12.34	15.12	16.98	19.81	22.36	25.47	27.69
14	13.34	16.22	18.15	21.06	23.68	26.87	29.14
15	14.34	17.32	19.31	22.31	25.00	28.26	30.58
16	15.34	18.42	20.46	23.54	26.30	29.63	32.00
17	16.34	19.51	21.62	24.77	27.59	31.00	33.41
18	17.34	20.60	22.76	25.99	28.87	32.35	34.80
19	18.34	21.69	23.90	27.20	30.14	33.69	36.19
20	19.34	22.78	25.04	28.41	31.41	35.02	37.57
21	20.34	23.86	26.17	29.62	32.67	36.34	38.93
22	21.34	24.94	27.30	30.81	33.92	37.66	40.29
23	22.34	26.02	28.43	32.01	35.17	38.97	41.64
24	23.34	27.10	29.55	33.20	36.42	40.27	42.98
25	24.34	28.17	30.68	34.38	37.65	41.57	44.31
26	25.34	29.25	31.80	35.56	38.88	42.86	45.64
27	26.34	30.32	32.91	36.74	40.11	44.14	46.96
28	27.34	31.39	34.03	37.92	41.34	45.42	48.28
29	28.34	32.46	35.14	39.09	42.56	46.69	49.59
30	29.34	33.53	36.25	40.26	43.77	47.96	50.89

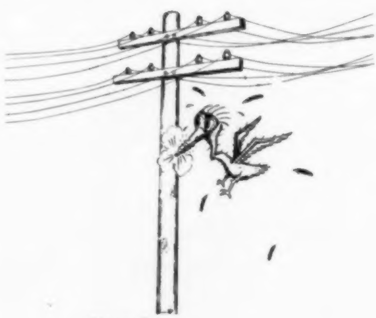
Table B is abridged from Table IV of Fisher and Yates: Statistical Tables for Biological, Agricultural, and Medical Research, published by Oliver and Boyd Ltd. Edinburgh, by permission of the authors and publishers.

ULTRASONICS - THEORY AND PRACTICE

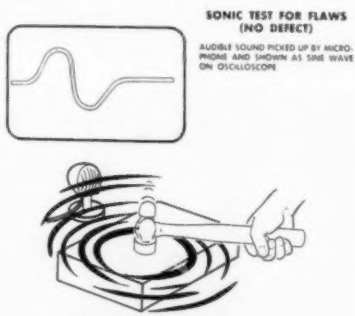
W. C. Hitt
Asst. Chief, Quality Control
Douglas Aircraft Co., Inc.
Santa Monica Division

The woodpecker is a rather clever bird, he knows how to make a sonic test to locate his meals. The shallow hole pattern left on a tree or pole are really test holes, indicating sound areas determined by resonance response or audible signals. Areas containing grubs or insects are located by the lack of resonance response or change in the audible response to his sonic tapping.

This method of testing has been used for many years in a manner such as the tapping of a metallic part and listening to the audible ring, indicating the part's soundness, as we know a cracked part will not produce the same sound response as would a sound part.



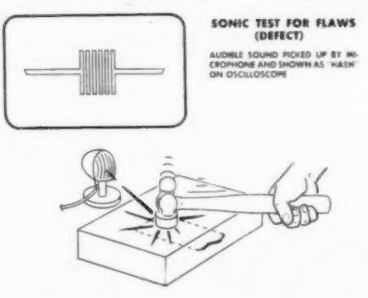
View 1.



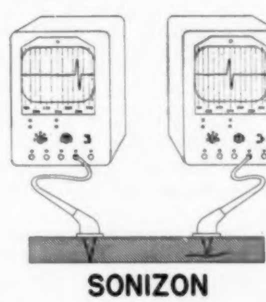
View 2.

With the aid of a hammer, microphone, and a cathode ray oscilloscope, it was possible to pick up these sound waves and picture them as a conventional sine wave pattern. (View 2.)

An unsound part whose discontinuities disrupt the propagation of sound waves may produce an unsymmetrical vibration pattern. (View 3.)



View 3.

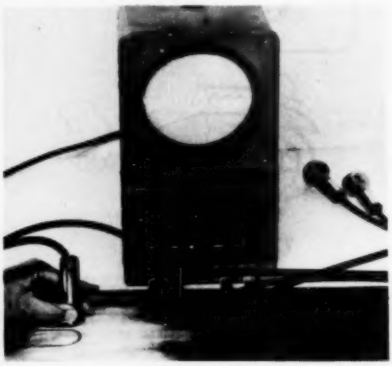


View 4.

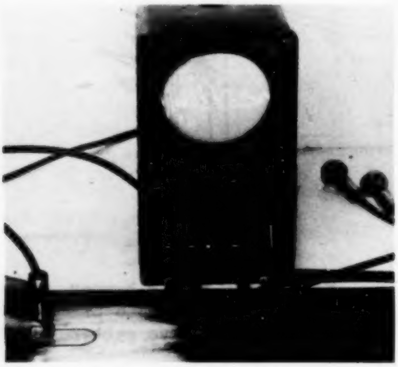
Ultrasonic testing as used by industry today falls into two general categories: Resonance and Pulse Echo. Resonance equipment such as the Sonizon is used for thickness measurement when only one side of the part is accessible, or when mechanical means are impractical due to the size and/or shape of the part.

As seen in View 4, left sketch, the trace on the oscilloscope is the point at which the vibrating crystal and the material are in resonance. This is in effect, a measure of the material density and thickness. A thinner part in this frequency range would produce a trace deflection on the left side of the screen. This may be seen in the right sketch which indicates the thickness of the material between the crystal and the lamination. The lack of bond associated with brazed joints and laminar type flaws illustrated in this view are found readily by this method.

In testing the sound area of an aluminum plate, you see the correct number of harmonic deflections of the trace, indicating by harmonic reflections, the plate thickness. In this case the thickness of the material is beyond the fundamental range of the machine and the oscilloscope shows multiple harmonics of the frequency range used. By using harmonics, it is possible to extend greatly the useful limit of the equipment. (View 5.)



View 5.

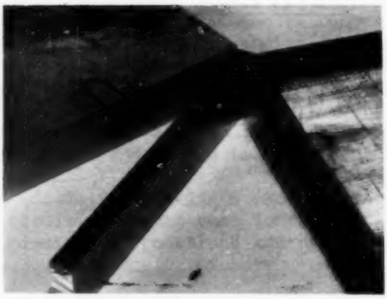


View 6.

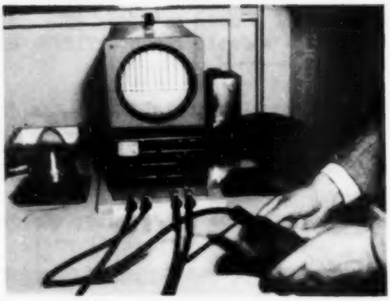
The lamination in this plate is shown by the change in the harmonic pattern seen on the oscilloscope (view 6.). As the dimension to the flaw or discontinuity is at the midway point, only half the thickness is indicated.

This is the actual flaw found by this test (view 7.). Its dimensions are $1\frac{1}{2}$" long by $\frac{1}{2}$" wide near the center of the plate. This was one of the first parts found defective by ultrasonic inspection at the Douglas Santa Monica Plant.

The resonance type equipment is best suited for wall gauge or thickness evaluation. Formed sections where thinning may occur, spun parts, castings, and exhaust pipes are excellent applications for this equipment. (View 8.)



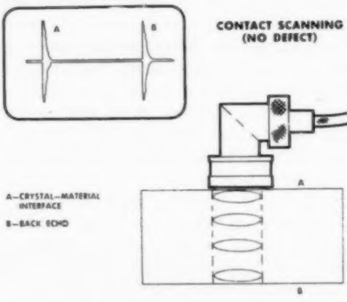
View 7.



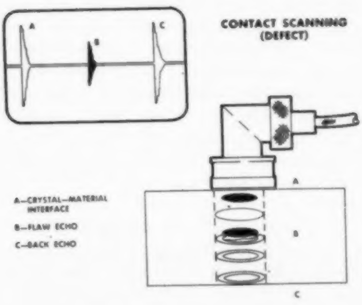
View 8.

The pulse echo method of flaw detection by contact scanning is illustrated in View 9. The top left view shows signal "A" indicating the crystal and part interface. The signal "B" indicates a reflected echo from the back of the part being tested, the line between "A" and "B" is the time base line. The right lower view shows the propagation of sound waves through the material being transmitted and received by the crystal.

The basic difference in this method and the previously described resonance method is that instead of reading the point of resonance, we measure the time it takes for the sound to travel through the part and return to the surface. Any flaw echoes would be indicated between these two points.



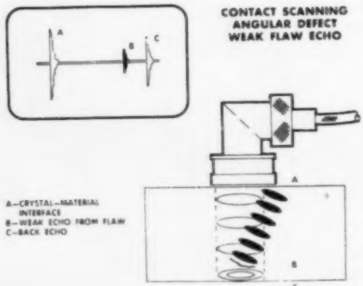
View 9.



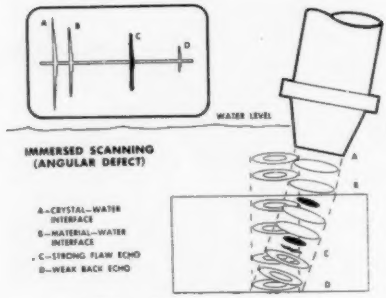
View 10.

The flaw echo "B" appears between the face and back echoes "A" and "C". The white pressure waves indicate the transmitted sound. The dark shows reflected echoes returning to the crystal as echoes from a flaw which lies near the center of the part. The plane of the simulated flaw is parallel to the face and back surfaces of the part. (View 10.)

We see in View 11 one of the shortcomings of contact scanning. In this case the plane of the flaw lies at an angle to the plane of the part. The reflected echoes are returning at the same angle as the plane of the flaw; a very weak echo is received as is noted in the view at the upper left. The flaw echo is identified with the letter "B". If we could move the crystal to the right and angle it to parallel the flaw echo, we could determine more accurately by the signal height, the cross sectional area of the flaw. In this application loss of back echo signal strength should be evaluated.



View 11.



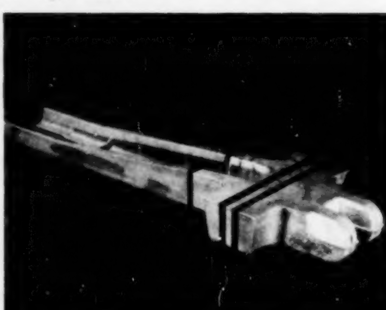
View 12.

This problem is solved readily by immersed testing when the crystal can be manipulated at will and lined up with the returning sound waves. When parts are inspected by the immersion method, an additional signal which represents the material and water interface is seen. In the upper left of View 12, signal "A" is from the crystal water interface; signal "B" is from the water material interface; signal "C" is from the flaw; and signal "D" is a weak signal from the back surface of the material. The flaw signal in this case indicated more accurately by its height the cross sectional area of the flaw.

The contact method of ultrasonic scanning presents several problems that must be considered and evaluated. It is limited to frequencies up to 10 MC, which in turn limits the method to parts of heavier cross sections. Thinner parts can be inspected at the higher frequencies used in immersed testing. The surfaces of the part being tested must be relatively smooth, which may require that the part be machined or sanded prior to testing. The one limitation of contact scanning that concerns us most is the loss of flaw signal strength when the flaw is not parallel to the surface of the part. Curved surfaces cannot be inspected satisfactorily.

Crystal wear in contact scanning causes the crystal to rock and gives the operator considerable trouble. The cost of replacing worn, cracked, and chipped crystals is considerable. This is particularly true of the thin high frequency crystals. Interface hash caused by surface roughness hides defects and confuses the operator.

View 13.



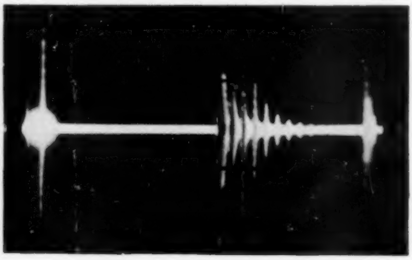
View 14.

The contact method of ultrasonic inspection is shown in View 13. The operator has coupled the searching crystal to the part with a light coating of oil. The trace displayed on the oscilloscope indicates either sound or faulty material.

View 14 shows a heavy forging that was found defective in ultrasonic inspection. A wafer taken from the thick section of the forging has been machined and etched to reveal the hidden flaws.

This black light photograph of the wafer indicates that the discontinuities do not come to the surface at any point which accounts for them passing the penetrant and etch tests. (View 15.)

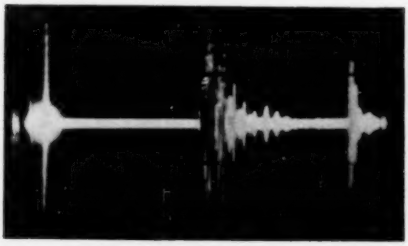
View 15.



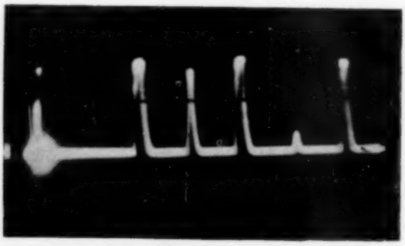
View 16.

Immersed scanning utilizes in practical production testing frequencies up to 25 MC. The advantage of high frequency testing is illustrated in Views 16, 17, 18, and 19. A conventional 5 MC immersed test of 5/8" thick plate with several multiple back reflections is seen on the oscilloscope. The working distance between the front and back echoes leaves a very narrow band to detect flaws. This picture is of good material, no flaw echoes are seen. (View 16.)

The hash seen between the face and back echoes shows a laminar type discontinuity. If the distance from the face of the material to the flaw is not critical, this type of flaw indication may be all that is needed. (View 17.)



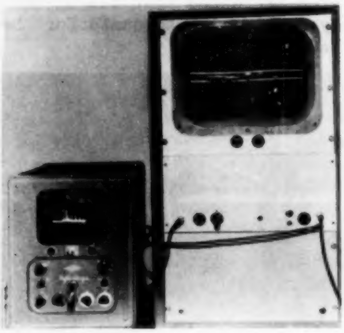
View 17.



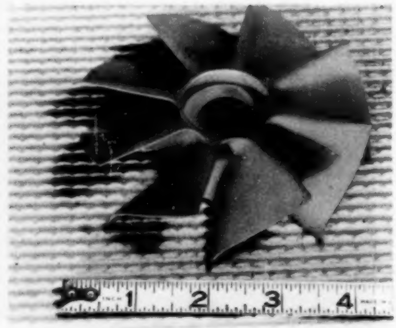
View 18.

The same discontinuity is seen between the face and back echoes in View 18. In this case 25 MC frequency was used with a video type "A" scan presentation. The wide band converter attached to the Reflectoscope is used in this immersion test.

The ideal presentation of the same piece of 5/8" laminated plate is seen in the large "B" scan or cross sectional view, (View 19.) This method lends itself to fast automatic scanning of plate, bar, and hand forging of regular shape. In this application the Reflectoscope is coupled to the wide band converter and "B" scan. The top line or trace on the large tube is the face of the plate; the lower broken line is the bottom surface of the plate; the line between the two indicates the lamination. The bottom line is broken due to reflections from the lamination masking, or shadowing out the back echoes.



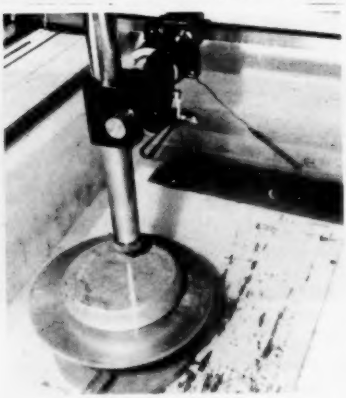
View 19.



View 20.

The machining of this part (View 20.) starting with a hand forging costs something like $80.00 each. One order of these forged blanks contained an excessive amount of discontinuities, which were found in the finished part by penetrant inspection. This led to the following ultrasonic testing.

An air driven turntable rotates the hand forging which the operator scans from the center to the outer rim. The ultrasonic testing time per forging is something less than five minutes each. (View 21.)



View 21.

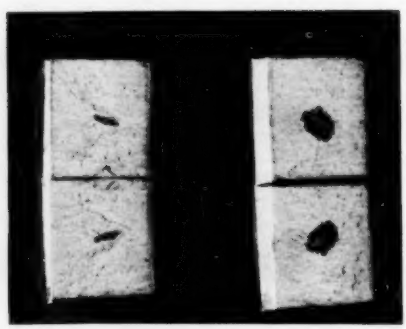
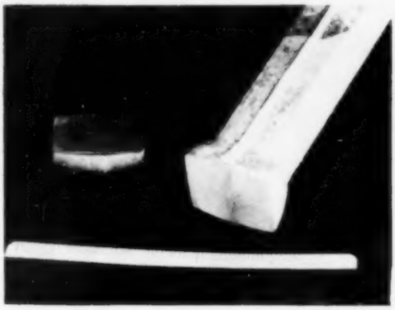


Figure 22.

A section of the forging is broken open at the defect (View 22.). The ultrasonic flaw signal height for the larger flaw was 1-3/4", while the smaller one registered a signal of 3/4" amplitude. This is really pin-pointing flaws, which could not be found by other non-destructive production testing methods.

The following pictures will show you typical flaws found by ultrasonic testing:

View 23. Forging cracks found in a 75ST aluminum control surface forging. The end has been sectioned and etched to reveal the forging defects.

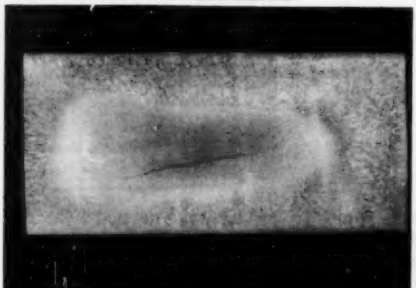


View 23.

View 24.

View 24. The defects seen in this picture are of a partially healed crack in an extruded piece of bar stock. The defects probably originated with a crack in the ingot used in extruding the bar.

View 25. A section removed from a hand forged block shows a subsurface crack approximately three inches long. As this flaw was close to the edge of the part and on an angle to the surface, the flaw was missed in contact scanning and found by the immersion test.

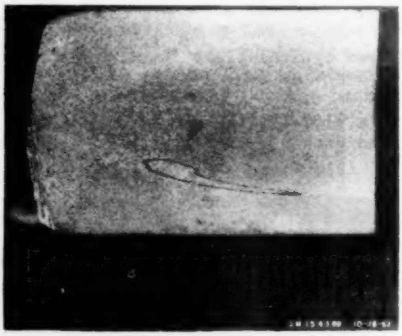


View 25.

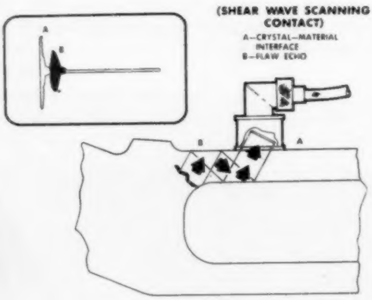
View 26.

View 26. This condition is known to us as a root end defect, a ring type laminar discontinuity at the re-crystallized grain boundary of the extrusion. It can be seen visually after etching and is found quite readily by ultrasonic testing without cutting and etching.

An odd shaped forging discontinuity in a die forging found by ultrasonic testing, the etch shows the area of flaw which is about $1\frac{1}{2}$" long. (View 27.)



View 27.

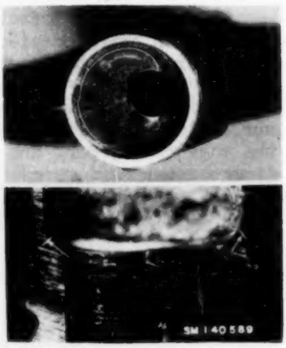


View 28.

The contact shear wave technique is used in many cases to locate cracks in parts that are not adaptable to immersion equipment. The crystal is mounted in plexiglass at an angle of 45° to the plane of the part. The path of the sound wave is seen in View 28.

View 29 shows the operator testing a closed assembly for internal cracks, the suspected crack is on the ID of the forged end, which is flash welded to the tube.

View 29.

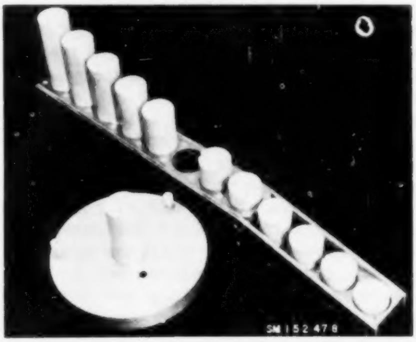


View 30.

A cross section of a cracked part found by shear wave testing at 5 MC. The fluorescent material used to reveal this crack after sectioning is Magnaglow. The evaluation of defects by ultrasonic testing, such as cracks, bursts, laminations, inclusions, and singular discontinuities are generally not too difficult. The cross sectional area of singular flaws can be estimated; the length, width, angle, and depth of larger flaws can be measured. One of the most difficult problems is the evaluation of multiple indications of porosity and stringer type discontinuities. (View 30.)

For singular type flaw evaluation, variable length test blocks having a 5/64" diameter flat bottom test holes are used. (View 31.) The distance from the top of the material to the defect in the part being tested is read on the oscilloscope. A test block of the same dimension is selected and the signal height is reset to 1" high. The defect in the part is subject to rejection on the basis of having a discontinuity larger than the cross sectional area of a 5/64" diameter flat bottom hole.

There is still considerable work left to be done in establishing standards. The Airframe Committee of the Society for Non-Destructive Testing is active on this work along with the A.I.A., A.S.T.M., Alcoa, and independent laboratories.

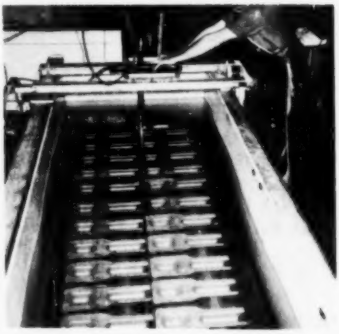


View 31.

View 32 shows a Douglas installation consisting of an electrically driven scanner in use in our Receiving Inspection Department for the inspection of raw forgings, plate, bar, and extruded materials.

The Reflectoscope, wide band converter, and "B" scan are used individually or collectively, depending on the parts being tested. These units present a picture of the internal structure of the material.

We presently are inspecting by ultrasonic methods, forgings and certain parts made from plate, bar, and extrusion, depending upon the end use or history of the part. We find this method of quality control assurance to be an economical and practical way of guaranteeing the part to be free from harmful discontinuities. The labor and material savings made possible by these tests more than offset the cost of inspection.



View 32.

View 33.

Ultrasonic testing offers a tremendous potential as a tool of quality control. Its application and capabilities are measured in feet rather than in inches of penetration. Its resolution, which may reveal a flaw the size of a pin head at several feet in ferrous or non-ferrous alloys, opens a new field of material evaluation. Ultrasonic applications complement rather than compete with any other method of flaw detection.

Photographs - Courtesy of Douglas Aircraft Co., Santa Monica, California
Sonizon - Magnaflux Corporation, Chicago, Illinois (View 5 and 6)
Reflectoscope - Sperry Products, Inc., Danbury, Connecticut (View 13)
Wide Band Converter and B Scan - Electro Circuits, Inc., Pasadena, California
Standards Blocks - Ultrasonic Testing & Research Laboratory, Van Nuys, California (View 31)
Immersion Equipment - Douglas Aircraft Co., Santa Monica, California (View 32 and 33)

HARDNESS AND ITS MEASUREMENT

Vincent E. Lysaght
Wilson Mechanical Instrument Division
American Chain and Cable Company, Inc., New York

It will be of interest to members of the American Society for Quality Control to know that one of two new applications in hardness testing is the use of statistical analysis to reduce the amount of hardness testing and still maintain good quality control of the product The second new application is the use of the microhardness tester which will be described later in this paper.

The application of statistical analysis is particularly suited to inspection of small parts which are produced in very large quantities. By proper application of statistical methods, and study and analysis of results, it is possible to control the various batches by performing hardness tests on only a very small percentage of the total production.

It is not the purpose of this paper to discuss the mechanics of quality control other than to mention that control charts may be used to assist in the maintenance of the calibration of hardness testing equipment and to test incoming material for hardness by use of a variable sampling plan, as well as being used in controlling the hardness of the manufacturer's own product. The purpose of this paper is to acquaint the quality control engineer with the practical side of hardness testing and how to make the best use of the hardness test.

Hardness as considered in this discussion is resistance to permanent deformation by an indenter of specific size and shape under a known load. This definition is the one most commonly applied to hardness testing from a metallurgical standpoint and it eliminates such concepts as scratch hardness, dynamic hardness, boring hardness, grinding hardness and abrasive hardness. Because of these various separate ideas of hardness no strict definition exists of the idea of hardness.

Much valuable information is obtained from a hardness test due to the relationship between hardness and other physical properties of metals. The hardness test is used for inspection and control purposes to make certain that the metal is brought to the best condition of heat treatment or cold work, or a combination of the two, for its particular use. In addition, the tensile strength of some materials may be estimated accurately for practical purposes from the indentation hardness number. Surface conditions may also be controlled.

In any consideration of the use of hardness tests it must be kept in mind that different companies utilize the hardness test for different reasons. Some use the hardness test to meet specifications; others for process control. Some are interested in hardness to the extent that there is a reliable correlation with tensile values - others want to predict fabricating properties. Many use the hardness test for quality control and others use it for determining the hardness of a material as a fundamental property of material.

DESCRIPTION OF HARDNESS TESTING EQUIPMENT

The commonly used testing machines for determining resistance to permanent indentation are the Brinell, Rockwell and Diamond Pyramid

hardness testers. These are known as indentation hardness testers.

The Brinell method was introduced in 1900 by Dr. J. A. Brinell and the hardness number is the ratio of the applied load in kilograms to the surface area of the indentation. The tester (Fig. 1) generally consists of a hand-operated hydraulic press designed to force an indenter into the specimen. The test is made with a 10-mm diameter ball under a load of 3000 kg. for ferrous metals and a load of 500 kg. for non-ferrous metals. The load is usually applied for 30 seconds and upon removing the load the diameter of the impression produced is measured with a microscope containing a graduated scale usually to 0.1 of a millimeter. The hardness number may be taken directly from a table.

The Rockwell* tester (Fig. 2) was invented in 1921 by S. P. Rockwell. In this test both a steel ball and diamond cone penetrator are used. A minor load of 10 kg. is applied first, after which a dial gauge is set at zero. The major load is added and removed. The dial gauge then reads the Rockwell hardness number directly. Thus the Rockwell test measures the additional depth to which a cone or ball is driven by a heavy (major) load beyond the depth to which the same penetrator has been driven by a definite light (minor) load. Several different Rockwell scales are used to cover the range of hardness of different metals. Each scale is designated by a separate letter which is associated with a particular major load and penetrator. The major loads are 60, 100 and 150 kg. and the penetrators, the 120 degree diamond cone penetrator, with 0.2-mm radius truly tangent to the cone, known as the BRALE* penetrator, and ball penetrators of 1/16", 1/8", 1/4" and 1/2" diameter.

The combination of diamond BRALE penetrator and 150 kg. major load is known as the "C" scale. It is used for testing all hardened and heat treated steel. The "B" scale is the combination of the 1/16" steel ball penetrator and 100 kg. load. It is used for testing unhardened steel, brass, bronze, etc.

The Rockwell superficial tester (Fig. 3), a specialized form of the Rockwell tester, operates on the same principle as the normal model Rockwell tester, but the minor load is reduced to 3 kg. and the major loads to 15, 30 and 45 kg. The 1/16" ball penetrator is used on brass, bronze and unhardened steel. On hard steel surfaces a diamond BRALE penetrator is used, but because of the smaller penetration, the radius of the penetrator must be ground with still greater micrometric precision and is designated as the "N" BRALE penetrator.

The Rockwell superficial tester has its own scales and these are designated first by the value in kilograms of the major load used; second by the letter "N", if the "N" BRALE penetrator is used, or the letter "T" if the 1/16" ball is the penetrator, and third the dial reading. The Rockwell superficial tester is used for testing thin materials and thin superficially hardened materials.

The 136 degree diamond pyramid hardness test, commonly known as the Vickers test (Fig. 4) follows the Brinell principle in that an indenter of definite shape is pressed into the material to be tested under a selected load, the load removed, the resulting impression measured and the hardness calculated by dividing the load by the surface area of the indentation. Loads up to 50 kg. are generally used and the indenter is

*Trade Mark Registered

Fig. 1 - The Brinell Tester

Fig. 2 - The Rockwell Hardness Tester

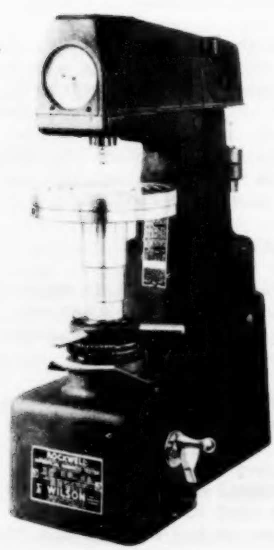


Fig. 3 - The Rockwell Superficial Hardness Tester

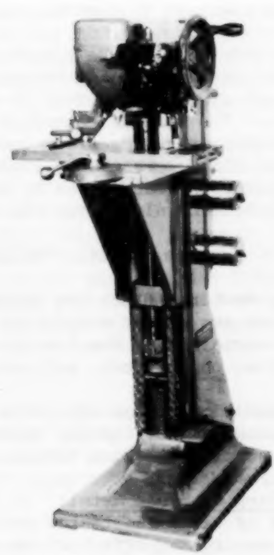


Fig. 4 - The Vickers Tester

a sharp-pointed square base diamond pyramid having an included angle between opposite faces of 136 degrees. The average length of the two diagonals is measured with a microscope and the hardness calculated from formula or by reference to a table. This test is used for research work and on sections of odd shapes which may be mounted in bakelite or clamped for testing.

Although the Scleroscope is a dynamic form of hardness tester it will be described here as it is oftentimes used for testing large pieces, especially where the instrument may be brought to the piece being tested. It was invented by A. F. Shore in 1907 and operates on the drop and rebound principle. A diamond faced hammer is dropped within a glass tube on the surface of the specimen being tested. The height of rebound on an arbitrarily marked scale is an indication of the hardness.

This would be a good place to discuss portable hardness testers which are now available in a variety of designs for use on different applications. Generally, portable machines are not as accurate as the conventional bench type tester and will not give as accurate a determination of hardness, but if the part cannot be tested in a more accurate machine, due to its size and shape or conditions surrounding its location, then a test determined on a portable tester may be of considerable value. The danger lies in the use of portable machines because they are less expensive than bench type testers. Portable testers fall into two groups - one a gooseneck type which takes work up to a given thickness and tests inward from the edge of the piece a relatively short distance. The Ames, Webster, King and Riehle testers are in this group. The second class consists of machines which measure hardness even to the center of large sheets and large pieces. The Barcol Impressor, Poldi and Ernst are typical of this class. No one portable tester is designed to meet all the requirements referred to above.

The file test may be considered along with portable testers. The surface of the piece being tested is rubbed slowly but firmly with the sharp teeth of the file until it is determined whether or not the file will bite. The test is generally limited to untempered hardened parts or for exploring surfaces for decarburization. In an effort to rehabilitate blind persons the Timken Roller Bearing Co. check the hardness of roller bearing races by blind persons who can detect whether the file will bite or slide over the surface by sound and feel.

SPECIAL PRODUCTION APPLICATIONS

Many factors affect the indentation hardness of metallic materials. Among these are proper support of material as the testing loads are applied; normality of the test piece to load; rate of load application; time of load on specimen, and spacing of indentations.

In addition, three important phases of hardness testing must be watched closely i.e. testing sheet metal, testing cylindrical parts, and the use of conversion tables.

Testing of Sheet Metal: In both the Brinell and Rockwell tests the thickness of the piece tested should be such that no bulge or other marking showing the effect of the load appears on the side of the piece opposite the impression. Usually, if the thickness of the piece is ten times the depth of the impression, the results are not appreciably inaccurate.

The approximate depth of indentation of values obtained on the "C" scale of the Rockwell tester may be determined by subtracting the reading from 100 and multiplying the difference by 0.00008". This value does not include the depth of minor load indentation.

All Rockwell tests on sheet metal must be made on a single thickness of the material. The use of additional thicknesses of the same material does not give the same effect as a solid piece of the same total thickness as the combined pieces.

The hardness testing committee of the American Society for Testing Materials has completed a survey of the work done on limiting thickness of materials for Rockwell testing. It is entitled "Survey of Investigations of Effect of Specimen Thickness on Rockwell Tests," and appeared in the October 1953 issue of ASTM Bulletin. The ideal solution would be a table showing the limiting thickness of each hardness for the different scales of the Rockwell tester and other testing instruments.

Testing of Cylindrical Parts: A second cause of error is the testing of cylindrical parts. Due to the curvature of the specimen, the hardness value may be reduced. In the Brinell test this error may be kept to a reasonable value if the average of the two principal diameters of the indentation is used as the equivalent diameter, provided that the minimum radius of curvature is equal to, or greater than, five times the radius of the indenting ball.

In the Rockwell test the error due to radius of curvature may be of considerable error if the diameter of the material is less than 1/2 inch. Under such conditions the diameter of the piece should be specified as the results are comparative only for rounds of the same diameter. If a flat is prepared on the cylindrical surface, it should be prepared carefully so as not to change the hardness and the fact that the test was made on a prepared flat should be noted. Fig. 5 illustrates the corrections when using the BRALE penetrator with the Rockwell tester and Fig. 6 when the 1/16" ball penetrator is used.

Conversion of Hardness Scales: It has been pointed out that there are different hardness testing instruments to measure the hardness of different types of work. The Brinell test, for example, is used for testing cast iron and the Rockwell test for sheet metals, heat-treated steel, finished ferrous and non-ferrous material, etc. However, it is often desirable to convert from one hardness scale to another. No conversion is mathematically exact as different machines use different shaped penetrators and loads, with resulting different cold working of the material under test. Also, it has been found that the degree of previous cold working will affect the indentation value and, consequently, the conversion of scales.

The safest way to meet specifications for hardness is to test the material under the same condition of test as the specification calls for. Never use conversion from one hardness scale to another on thin material where the metal cannot be properly tested under both hardness scales. Use conversion on flat surfaces only.

Conversion of different hardness values is helpful in a general way, but should be used with discretion. The U. S. National Bureau of Standards, after a long investigation of this subject, determined the relation between Rockwell and Brinell numbers within an expected error

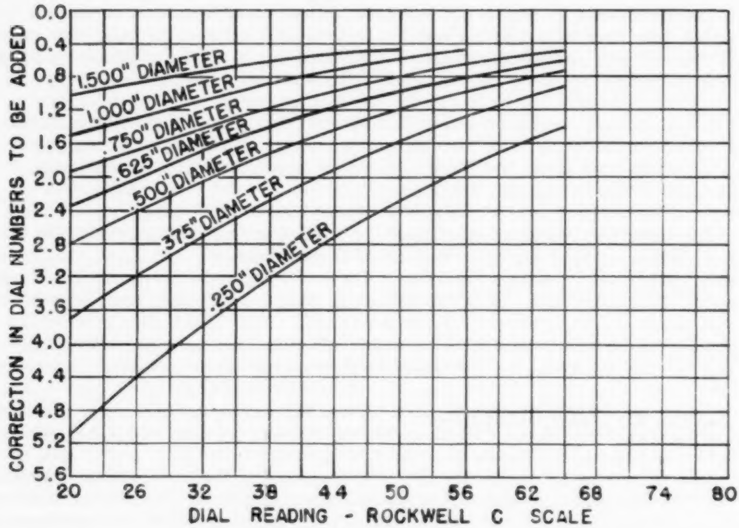


Fig. 5 - Hardness Correction Factors for Cylindrical Surfaces - Rockwell C Scale

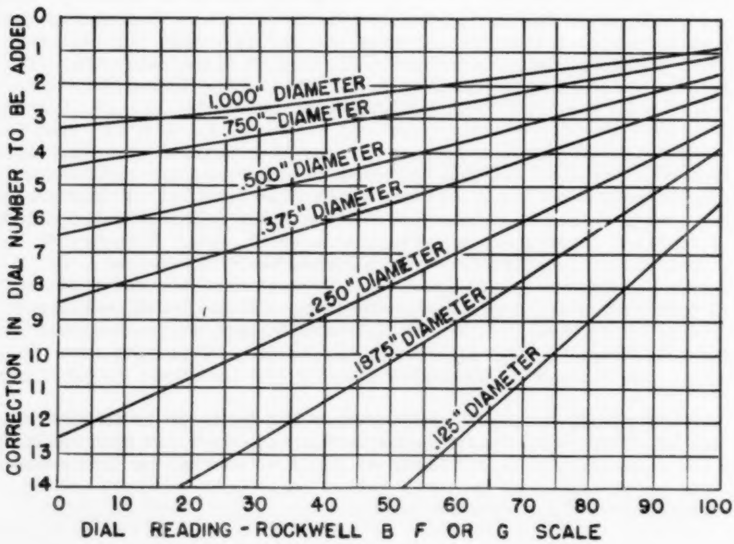


Fig. 6 - Hardness Correction Factors for Cylindrical Surfaces - Rockwell B, F or G Scale

of plus or minus 10%. Much work is now being done on conversion with the hope of preparing charts or tables which, when used under certain conditions, will permit good agreement. This work has not progressed sufficiently for general adoption as yet. Such factors as modulus of elasticity and work hardening capacity of the metal, have been found to contribute materially to conversion relationship. A universal conversion chart which has stood the test of time, and been in use since 1938, is shown in Fig. 7.

Irregular shaped pieces must be properly supported on specially designed fixtures if the hardness of the parts is to be maintained within predetermined limits in controlling quality.

To accomplish this end, special fixtures, penetrators, penetrator extensions, goosenecks and even special machines have been designed for testing irregular shapes. The tool engineer plays an important part in this work and a few illustrations will show the use of some of these fixtures. Wherever possible these fixtures should be designed to allow their use on several parts, in some cases even making the gage adjustable. However, it is not always possible or feasible to do this.

Fig. 8 illustrates an adjustable fixture, designed to hold gears in a position to check the hardness of gear teeth on the face of a tooth. The two holding blocks can be set at any position and the two rolls are held on the blocks by clamps, allowing the use of various size rolls. A special BRALE is used on this set-up. Note that the base of the fixture has a graduated scale used to locate the blocks. In practice a chart has been developed which shows the part number, the setting of the blocks, the size of the rolls to be used and the scale used on the Rockwell tester to be used in testing various gears.

Fig. 9 shows a universal fixture for testing gears ranging from small prism gears to maximum capacity of fixture for Rockwell tester.

For testing large and bulky pieces a special machine, designated as a Universal Testing Unit (Fig. 10) is recommended. In such conditions the unit is mounted on a special fixture designed to accommodate the large and bulky piece. The testing unit is lowered to the piece being tested rather than elevating the test piece on a spindle. If considerable work is to be tested the unit may be motorized.

Many hardness testing machines are used in production inspection. A typical example is the inspection of compressor blades forged for General Electric aircraft jet engines. The machines (Fig. 11), of both the automatic and the semi-automatic types, are operated by women on a two-shift basis. Each operator is able to check from 2000 to 2500 blades per shift. A minimum of two readings is taken on each blade, which is then dropped into one of three chutes according to hardness; classified as too hard, correct hardness, or too soft. Those not passing are re-cycled through heat treat and returned to the Rockwell stations again.

Fig. 12 shows a gooseneck plunger extension used for testing internal surfaces of cylindrical parts. The weight of this extension is taken into consideration when the machine is designed; otherwise incorrect loads could be applied - thus giving incorrect readings. These machines are semi-special.

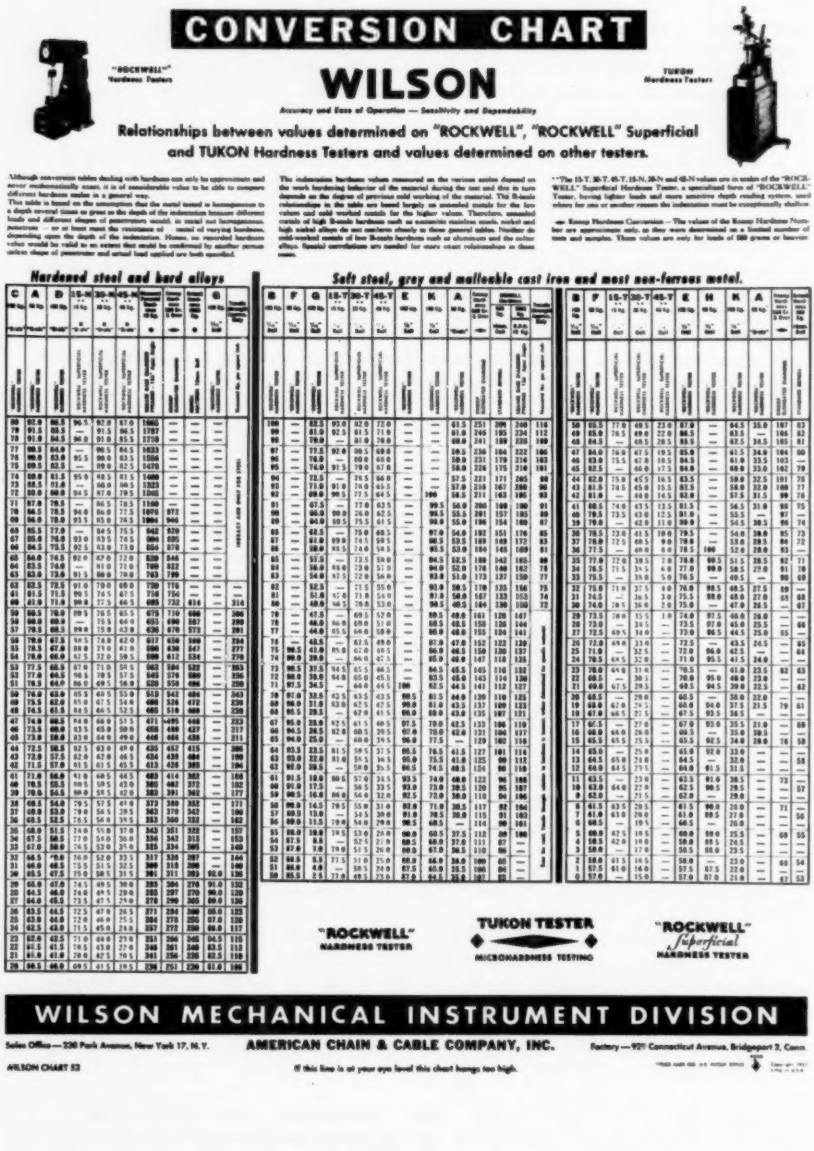


Fig. 7 - Conversion Chart

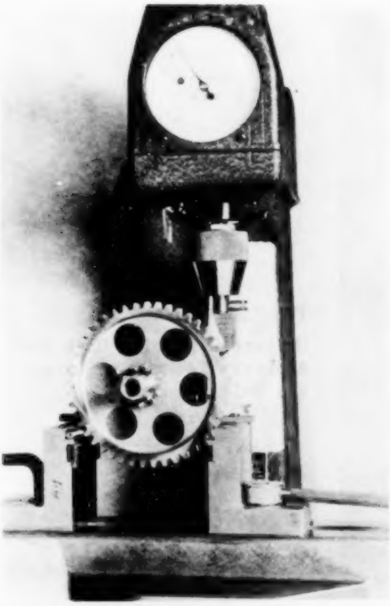


Fig. 8 - Rockwell Testing of Irregular Shaped Parts

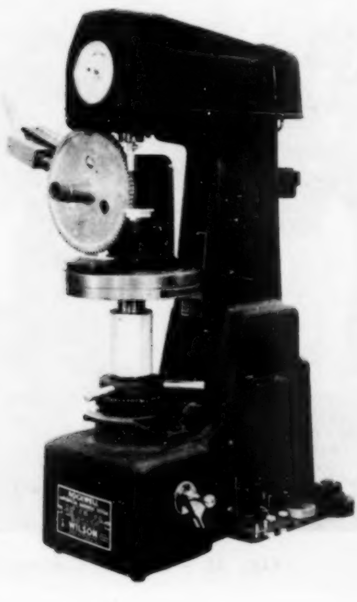


Fig. 9 - Universal Fixture for Testing Gears on Rockwell Hardness Tester

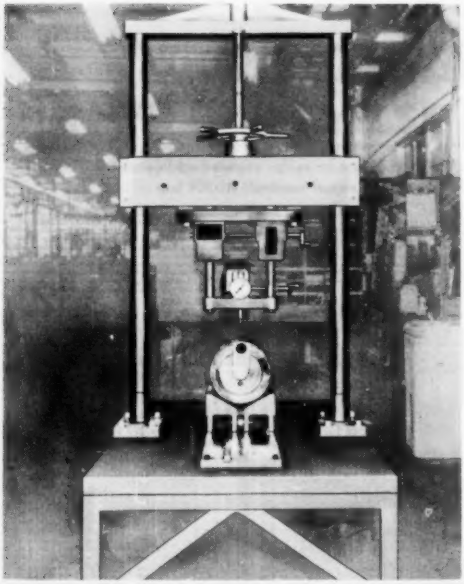


Fig. 10

Universal Testing Unit for Rockwell Testing of Large and Bulky Pieces

Fig. 11 - Testing Compressor Blades for Rockwell Hardness

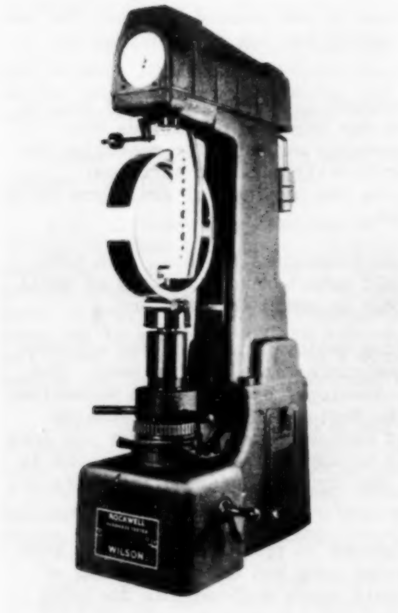


Fig. 12 - Testing Internal Surfaces with Gooseneck Plunger Extension

Fig. 13 - Tukon Microhardness Tester

At the beginning of this discussion it was pointed out that the use of the microhardness test was the second of two new applications of hardness testing.

The hardness testers we have discussed up to this point do not, as a general rule, give any indication of the hardness of various constituents in an alloy nor are they generally suitable for measuring the hardness of thin sheets or thin, superficially hardened surfaces; neither can they be used for controlling the hardness of small precision parts such as found in a watch movement.

Such testing is now done on the microhardness tester. This term, unfortunately, is misleading as it could mean hardness testing of small hardness values where it actually refers to small indentations.

The apparatus used for such testing employs low loads and accurate measurement of the indentation over restricted area must be made. The Tukon microhardness tester (Fig. 13), developed by the Wilson Mechanical Instrument Division and approved by the National Bureau of Standards after exhaustive tests, is one type of hardness tester used for low load testing. Loads as low as one gram may be applied although most work is carried out with loads of 100 grams. Two types of diamond indenters are used.

The Knoop indenter (Fig. 14) is ground to pyramidal form that produces a diamond shaped indentation having long and short diagonals of approximate ratio of 7 to 1. The pyramid shape employed has included longitudinal angles of 172° 30' and included transverse angle of 130° 0'. The depth of indentation is about 1/30 its length. It can be observed from the geometry of the indenter that indentations of accurately measurable length are obtained with light loads. The indenter was developed at the National Bureau of Standards. The Knoop hardness number is the load divided by the unrecovered projected area.

The other indenter used for microhardness testing is the 136 degree diamond pyramid indenter (Fig. 15) in the form of a square base with an angle of 136 degrees between faces. The depth of indentation is about 1/7 of the diagonal length. For certain types of investigations there are advantages to such a shape. The diamond pyramid hardness number is the load divided by the surface area of the indentation.

For making indentations in selected small areas, an accurately designed mechanical stage is necessary. An area of a few thousandths of a square millimeter can be accurately located under a metallurgical microscope built into the Tukon. The piece being tested is then moved under the indenter on the moving plate of the stage, the indentation made in the selected location, and the specimen returned under the microscope for the purpose of reading the dimension of the impression.

There are numerous types of microhardness testers available today in addition to those designed along the same general principle of the Tukon tester. Some of these have the indenter mounted in the front lens of an objective in a microscope. Others provide for the specimen to be attached to a counter-balanced beam. Most of these are in reality attachments to a bench or metallurgical type microscope. They

are either weight-loaded or spring-loaded and have many novel methods of applying and calibrating the load.

The method of preparing the surface of the material for testing may considerably affect the hardness value. Other factors affecting the results are the speed of loading and orientation of specimen with respect to crystallographic planes.

A few illustrations will show what may be accomplished by microhardness testing.

Fig. 16 shows microhardness of carbides in high speed steel with 136 degree diamond indenter. Magnification 2000 X. Hardness of carbide equivalent of Rc62. The use of the Knoop indenter to investigate the service failure of a 5/16 x 24 thread, 2 flute high speed steel tap, is shown in Fig. 17. The Knoop hardness traverse indicated that the heat developed in grinding and the subsequent quenching had formed hard and soft zones in the threads. In a similar manner all cutting tools can be studied for hardness at the extreme tip.

Small precision watch parts may be tested for hardness by mounting in thermosetting plastic and polished like metallographic specimens. Since the size of the indentation can be varied by varying the load, it is possible to successfully test every type of watch part, even the smallest. Frequent tests have been made, for example, on balance staff pivots, screws, pinions, studs, pins, clock springs, etc.

Instrument pivots, surgical needles and tiny pellets of pen points are among other small parts tested with the Knoop indenter.

The quantitative control of the hardness of shallow electro-plated surfaces or other hard, thin surface coatings, and the probing of limits of decarburization has for many years offered a challenge to the hardness test. The Knoop indenter has been a great aid in measuring the hardness of different electroplates. Although the work has not been confined to chromium, considerable work has been done on the determination of the hardness of the chromium plate.

With reference to the effect of various base metals on the hardness of chromium plate, one authority who has done considerable work along this line, reports that the base metal has no affect on the hardness of the plate provided the plate is of reasonable thickness and has satisfactorily measured hardness of plates .0005" thick without any apparent effect of the base metal. This figure confirms work done at the National Bureau of Standards.

Hardness of plated surfaces has been tested on the Tukon tester and impressions made on electrodeposits of cadmium, silver, zinc, copper, nickel and chromium are shown in Table I. These values should not be considered as representative of the plated surfaces under all conditions. Such factors as current density; temperature and composition of plating solution; valuation in hardness from the outside to the inside of the plate; structure of the electrodeposited metal, all influence the hardness.

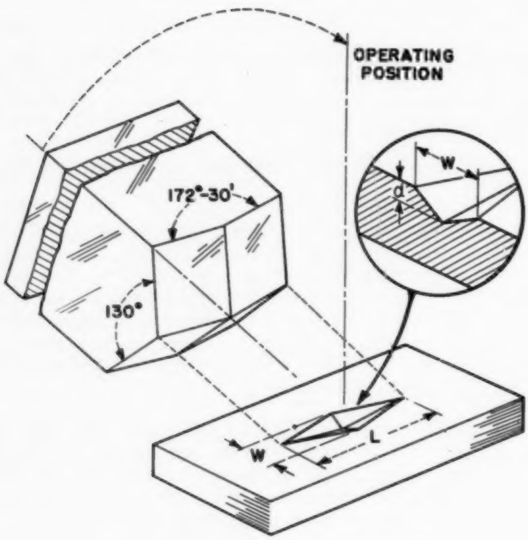


Fig. 14 - The Knoop Indenter

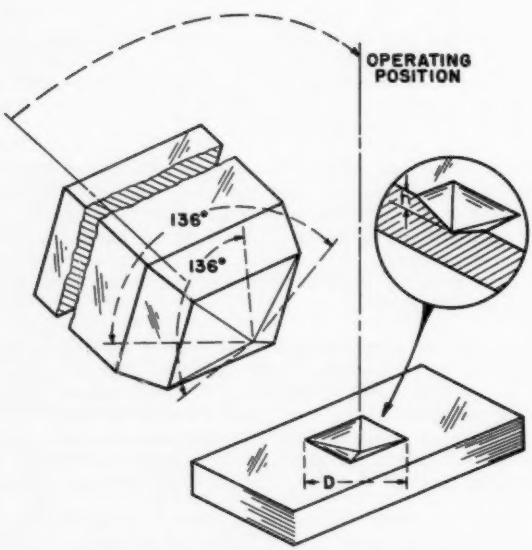


Fig. 15 - The 136 degree Diamond Pyramid Indenter

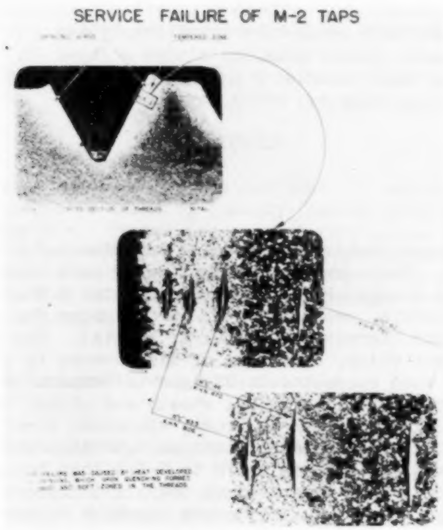


Fig. 16 - Microhardness of Carbides in High Speed Steel

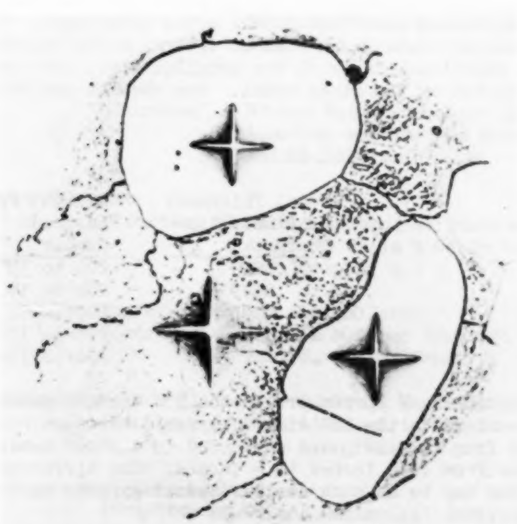


Fig. 17 - Knoop Microhardness Tests on a High Speed Steel Tap

TABLE I
Hardness of Electrodeposited Metals

Metal	Knoop hardness at 100 gram load
Cadmium	37
Silver	60
Zinc	120
Copper	165
Nickel	550
Chromium	935

The microhardness tester is now being introduced for production control of thin metal. One example will suffice to show what can be done in this field. The General Plate Division of Metal & Controls Corp., Attleboro, Massachusetts, use a Model FB Tukon tester for controlling the hardness of their Thermostat - Metal or Bimetal. The Bimetal is made up of two layers which, by virtue of differences in co-efficient of expansion, deflect when subjected to changes in temperature. The largest percentage of their material is nickel steels and nickel chromium steels of equal thickness. Alloys such as brass, stainless steel, chrome iron, silicon bronze and others, are also combined into thermostat metals. The metals are bonded together in 1-1/8 to 2-1/4 inch thick slugs and rolled into thin strip down to a minimum overall thickness of .003 inches commercially. Experimentally strip has been rolled to .0005".

The hardness test is an indication of the forming qualities and of the control of the yield point. It is a final check of the annealing and reduction processing cycle.

The 136 degree diamond pyramid indenter is used and the load is varied according to the thickness of the metal under test. Considerable laboratory testing was done to determine the suitable load to test only each individual layer of the metallic strip and not to have the results affected by the other metal. The results are shown in Table II below.

TABLE II

Load	Total Thickness of Bimetallic Strip	Diamond Pyramid Hardness (Approx.) Number
25 grams	.001 to .0029"	200 to 350
100 grams	.003 to .0049"	200 to 350
500 grams	.005 to .0095"	200 to 275
1000 grams	.005 to .0095"	275 to 350
2500 grams	over .010"	200 to 350

The tolerances are ≠ 25 numbers in the 350 diamond pyramid hardness range to ≠ 15 numbers in the 200 diamond pyramid hardness range. Small samples are cut from the strip and supported in a sheet metal clamp. The width varies from .040 inches to 4 inches. The difference in hardness on each side may be as much as 100 diamond pyramid hardness numbers.

The microscope used in measuring the impression employs the 6-mm objective lens. The filar micrometer eyepiece has a magnification of 12.5X. Tables are prepared relating filar numbers to diamond pyramid hardness numbers.

The microhardness tester is housed in a telephone booth-like structure in the inspection department. Seven inspectors use the tester and have at least a week's training before being permitted to use the instrument. The engineering department also has access to the Tukon tester. Several hundred tests are made each week.

CONCLUSION

In closing it should be emphasized that the hardness test provides a quick and inexpensive means of quality control when hardness is an indication of quality. Proper use and interpretation of quality control methods will materially reduce the cost of inspection and still insure uniformity in hardness. The importance of the new art of microhardness testing for controlling the hardness of small parts should not be overlooked.

Photographs used in article were obtained from:

Fig. 1 - The Brinell Tester (Courtesy of Tinius Olsen Testing Machine Co.
Willow Grove, Penna.)

Fig. 2 - The Rockwell Hardness Tester
(Courtesy of Wilson Mechanical Instrument Division
American Chain & Cable Company, Inc.
New York, N. Y.)

Fig. 3 - The Rockwell Superficial Hardness Tester
(Courtesy of Wilson Mechanical Instrument Division
American Chain & Cable Company, Inc.
New York, N. Y.)

Fig. 4 - The Vickers Tester
(Courtesy of Riehle Testing Machine Division
American Machine and Metals, Inc.
Moline, Illinois)

Fig. 7 - Conversion Chart
(Courtesy of Wilson Mechanical Instrument Division
American Chain & Cable Company, Inc.
New York, N. Y.)

Fig. 8 - Rockwell Testing of Irregular Shaped Parts
(Courtesy of Pratt & Whitney Aircraft
East Hartford, Conn.)

Fig. 11 - Testing Compressor Blades for Rockwell Hardness
(Courtesy of General Electric Co.
West Lynn, Mass.)

Fig. 13 - The Tukon Microhardness Tester
(Courtesy of Wilson Mechanical Instrument Division
American Chain & Cable Company, Inc.
New York, N. Y.)

STATISTICAL TECHNIQUE FOR FORECASTING SALES AND THE LIASON WITH PURCHASING, SCHEDULING, PRODUCTION, SHIPPING, AND INVENTORY CONTROL

Guy G. Parkin
Minnesota Mining & Manufacturing Company

This paper is intended as a straightforward exposition of a concept whereby the mathematical approach of applying sound, scientific principles to sales forecasting and allied factors are tempered with entrepreneur's judgment, based on experience and intuition.

Sales forecasting apparently has a fantastic supernatural appeal, perhaps evolving as from a Houdini with the lamps of Aladdin.

I recently had occasion to present this subject matter at one of our Midwest section clinics. I was milling with the crowd in the corridor just prior to the opening of the clinic and overheard the conversation of a group relative to the session they would attend. One of the group remarked as follows, "You know, these S.Q.C. engineers sure go all the way. They have been and are predicting the quality capabilities of our industrial repetitive processing, and now, according to this sales forecasting paper, they are going to tell us how much we are going to sell. This I have to hear."

Not long ago I read an article by one of our better authors on "Statistics Applied to Industrial Processing" in which he made the following statement, "Industrial statistical control is comprised of approximately 80% sound engineering judgment and 20% mathematics." If there is a technique in our category of S.Q.C. concepts that conforms to this author's statement, it is the one we will discuss in this paper. The closest we will come to the supernatural is in our attempt to reduce or eliminate "ghost" sales and production.

The source of our interest in sales forecasting originated in our sales inventory department. When we were called in on this problem we found arithmetic sales levels (averages) estimated from sales history and tempered with management's survey of such factors as competition, obsolescence, advertising, national emergencies, etc. These levels were being predicted with a satisfactory degree of confidence. We also found that a coefficient of variation was recognized as present and inevitable but with no way of measuring it.

If we could evaluate the variability of these sales values around their satisfactory levels, we could contribute to the validity of maximum and minimum sales capability.

If a valid forecast of sales could be made, the following benefits would accrue:

1. We could contribute to the control of raw material purchasing, which would influence optimum investment, warehousing, and service to the production department. A predicted consumption of raw material contributes to good vendor relationship, particularly vendor quality.

2. We could contribute to the economy of scheduling production. We could eliminate the unsubstantiated fluctuations of production such as expediting "ghost" labor, material, and

facilities one month only to find these efforts nullified in the succeeding months. The smooth flow of a predicted product thru production is a major factor in economical production and good quality.

3. The capability limits of a valid sales forecast will establish factual, objective inventory values. Opinions and subjective guessing are eliminated. The costs of warehousing "ghost" merchandise for "ghost" sales are reflected in management's investment in subjective product inventory and warehousing space. If the product is perishable, quality becomes an economical factor.

4. Without factual sales process capability limits, chance inherent, sales variability may be mistaken for a systematic, chaotic condition. The resulting hysteria may contribute to unwarranted expenses in correspondence, telephone calls, telegrams, invoice writing, clerical, and administrative expense.

5. Valid sales prediction would contribute to economical transportation of the finished product from factory to sales distribution branch. Economic utilization of shipping facilities may be planned when shipping increments are predicted.

BASIC CONCEPTS OF SALES FORECASTING

To avoid excessive abstraction, we will expound this subject matter with a concrete example whose values are hypothetically chosen to illustrate procedures resulting from an invalid forecast.

Sales forecasting on a product which has been supplying a diversified market over a relatively long period of time is necessarily based on sales history. The arithmetic average of the sales history may be tempered with management's systematic plans, intuition, and judgment; however, the need of a tool indicating the occurrence of significant departures from the expected becomes apparent.

The further in advance of actual production this tool can perform its function, the more valuable it becomes relative to economical purchasing (raw material), scheduling, production, inventory, and transportation.

We have attempted to evaluate unsubstantiated fluctuations with a control technique designed to indicate automatically, invalid behavior relative to history. The control limits we use are approximately 95% confidence limits, however, the procedure we use in posting data (reserving an invalid prediction until two consecutive values, on the same side of the mean, fall outside a capability limit) allows us to infer with approximately 99% confidence. Assuming the factors involved are of a random nature, we can calculate the probable variation for individual months using statistical concepts. The mechanics of the technique are not difficult and can be operated by average clerical personnel.

TECHNIQUE

Referring to Chart I - The "Work" Form

Columns 1 and 2 - The history sales data (1951) from which we will forecast 1952 behavior.

Column 3 - Cumulative monthly sales (observed).

Column 4 - Average cumulative monthly sales (expected, if we had no variability).

Column 5 - Difference between expected and observed (monthly cumulative sales error).

Column 6 - The plus and minus errors are respectively squared, then added, divided by frequencies, and a square root taken. This value is the standard error of estimate, which is the root mean square deviation of the actual cumulative monthly values from the average cumulative monthly values and identifies the percent of random deviations that will fall within specific capability limits. One standard error will include 68.26% of all random deviation, two standard errors will include 95.46%, and three standard errors will include 99.73% of all random deviation.

The hypothetical values used in Chart I reveal a mean ($\overline{X}$) = 9.4, and a standard error of estimate (σ_{λ}) = 3.0.

Referring to Chart II - The "Control" Form in which we depict the "News Reel" validity of 1952 sales behavior and liason with inventory.

From Chart I (Type 560), 1951 sales history, we have obtained an average monthly sale $\overline{X}$ of 9.4 and a standard error (σ_{λ}) of 3.0. In designing Chart II we will assume that the sales of (Type 560) behave in 1952 as they did in 1951, and accordingly, compute sales capability limits (column 7) within which random 1952 values will be found.

Column 1 - Identifies the 1952 months.

Column 2 - Identifies the net inventory (obtained by subtracting actual sales, column 5, from gross inventory, column 4).

Column 3 - Identifies the average predicted shipping increment (the $\overline{X}$ of history values).

Column 4 identifies the gross inventory (net inventory plus shipped).

Column 5 - Identifies the actual monthly 1952 sales.

Column 6 - Identifies the cumulative actual monthly 1952 sales.

Column 7 - Identifies the capability limits, predicted from 1951 behavior, within which 1952 monthly cumulative sales will fall if onlv random and not systematic influences are present.

ILLUSTRATION OF THE MECHANICS OF THE OPERATION OF CHART II

Using hypothetical sales values for 1952 (column 5), we will test the validity of the forecast made for 1952 from 1951 sales. First, we will compute 1952 cumulative sales capability limits (column 7) for each month of 1952 based on 1951 behavior using $\bar{X} = 9.4$ and $\sigma_A = 3.0$. These limits are set with approximately 95% confidence (the cumulative mean $\pm 2\sigma_A$). Secondly, we will protect against maximum (99.73%) probability of sales for January ($\bar{X} + 3\sigma_A$ or $9.4 + (3 \times 3) = 18.4$) and enter this value in the first row in column 4.

With these preliminary entries we are ready to receive the monthly actual sales values, record them in column 5, post cumulative sales in column 6, and verify the validity of our forecast by checking the cumulative sales value against the maximum and minimum capability limits (column 8) for that specific entry. If the cumulative sales entry is outside its capability limits ($\bar{X} \pm 2\sigma_A$). We infer with 95.46% confidence that the sales process has sought a new level and our prediction of 1952 sales behavior being the same as 1951 is invalid. If the cumulative sales value is found within its capability limits, our prediction remains valid.

You will note in the case history Chart II that the April cumulative sales entry was out of its ($2\sigma_A$) capability limits. Inasmuch as there are 5 chances in 100 of our being wrong by inferring the prediction as invalid at this time, we reduce this probability of error by giving the sales process another opportunity to portray its behavior. The $\pm 2\sigma_A$ used will normally allow values to fall outside its limits by pure chance on the same side of the mean, for two consecutive months only about 1% of the time. If the May entry, of the case being studied, is out of its $2\sigma_A$ capability limits, the probability of error in inferring invalid prediction is one in 100. The May entry is out of its control limits and we assume our prediction as not valid and proceed to find the statistics $\bar{X}$ and σ_A of the new sales process.

Referring to Chart II, columns 1 thru 5 and beginning with the 18.4 gross inventory, you will note the simplicity of calculation required by the various column entries. The training qualifications required for personnel making these routine entries is of small magnitude.

You will note (Chart II) during the period of January thru April when we first noted some anomaly, that our shipping increment has been constant. During this interval our net inventory has been decreasing. In order to protect the process against the probability of a "short" gross inventory in May, the May shipment was increased by an amount sufficient to build the May gross inventory back to the January original 18.4. The out-of-control warning in column 6 is the signal for net inventory investigation (too large or too small).

Inasmuch as our forecast for Type 560, Chart II, had been found invalid in May 1952, we proceed to seek this new level and standard error of estimate and predict future behavior on the strength of this new finding.

Where adequate history is not available we find it necessary to investigate all factors that might have had or may have a probability of contributing to a changed level and temper our inferences accordingly.

In our case, Type 560, we found our estimate of the true new level to be 11.5.

In our studies involving factors of a random nature, we have found orders from a customer varying in size from year to year relative to general business conditions; however, the relationship of the σ to the $\overline{X}$ (the coefficient of variation) is relatively constant, therefore a coefficient of variation $\sigma/\overline{X}$ for a specific class has been used to evaluate the σ for a new level. Chart II, Type 560, upper right hand corner disclosed an original coefficient of variation $\sigma/\overline{X}$ or 3/9.4 = 31.9%. In determining our new σ from the new level 11.5, we find the new σ = 31.9% of 11.5, or 3.7.

With a new $\overline{X}$ = 11.5 and a new σ =3.7, we are ready to forecast the future behavior of Type 560 and proceed to establish new monthly cumulative sales capability limits (column 7) using the same logic as previously.

A study of Chart II will reveal cumulative sales values for the balance of 1952, within their respective predicted capability limits. The prediction made in June 1952 for sales behavior was valid thru December 1952. The shipping has remained constant at 11.5. The net and gross inventories have remained at a satisfactory level.

You will note that in designing Chart II, the "Control" Form, we have avoided the probability of frightening a layman posting clerk with a slope line control chart. We have found the accessibility and sensitivity of column 7 to be extremely practical.

The "Control" Form, Chart II, was designed to display the relationship between sales, gross and net inventory. If we could predict sales, we could set up a predicted "chain" reaction affecting the prediction of inventories, warehousing, transportation, scheduling, production, purchasing, and the multitude of clerical and administrative functions connected with these factors. Valid sales forecasting could reduce biased, opinionated, subjective guessing relative to the moving targets of "ghost" inferences.

CHART 1

Jan 1, 1952 TYPE - 560	SALES ACTUAL	(X) CUMULATIVE ACTUAL	(X') CUMULATIVE AVERAGE	(X)-(X')	$(x-x')^2$		
JAN. - 1951	17.3	17.3	9.34	+7.96	63.36		
FEB. - "	3.4	20.7	18.68	+2.02	4.08		
MAR. - "	6.4	27.1	28.02	-0.92	0.85		
APRIL - "	7.9	35.0	37.36	-2.36	5.57		
MAY - "	10.4	45.4	46.70	-1.30	1.69		
JUNE - "	10.0	55.4	56.04	-0.64	0.41		
JULY - "	7.8	63.2	65.38	-2.18	4.75		
AUG. - "	12.4	75.6	74.72	+0.88	0.77		
SEPT. - "	10.9	86.5	84.06	+2.44	5.95		
OCT. - "	9.2	95.7	93.40	+2.30	5.79		
NOV. - "	9.6	105.3	102.74	+2.56	6.55		
DEC. - "	6.8	12)112.1	112.08	—	11)99.27		
		9.34			9.024		
		$\bar{X} = 9.34$			$\sqrt{9.024} = 3.0 \rightarrow \sigma_d = 3.0$		

CHART 2

TYPE 560 IDENTITY	NET INV.	SHIPPED	GROSS INV	ACTUAL SALES	CUMULATIVE SALES	$\bar{X} \pm 2\sigma$ CAPABILITY – LIMITS –		
JAN. '52	—	—	* 18.4	11.4	11.4	3.4	15.4	* 9.4 + (3x3.0) = 18.4
FEB. "	7.0	9.4	16.4	10.6	22.0	12.8	24.8	
MAR. "	5.8	9.4	15.2	11.1	33.1	22.2	34.2	
APR. "	4.1	9.4	13.5	12.2	45.3 Δ	31.6	43.6	Out Control
MAY "	1.3	17.1	* 18.4	14.3	59.6 ΔΔ	41.0	53.0	2nd Out Control * Adjust to 18.4
JUNE "	4.1	18.5	* 22.6	12.8	12.8	4.1	18.9	* 11.5 + (3x3.7) = 22.6
JULY "	9.8	11.5	21.3	11.9	24.7	15.6	30.4	
AUG. "	9.4	11.5	20.9	11.0	35.7	27.1	41.9	
SEPT. "	9.9	11.5	21.4	12.7	48.4	38.6	53.4	
OCT. "	8.7	11.5	20.2	13.7	62.1	50.1	64.9	
NOV. "	6.5	11.5	18.0	10.2	72.3	61.6	76.4	
DEC. "	7.8	11.5	19.3	11.4	83.7	73.1	87.9	

DATE	$\bar{X}$	σ	2σ	3σ	Coef. Var.
1/52	9.4	3.0	6.0	9.0	31.9%
1/52-5/53	11.5	3.7	7.4	11.1	11.5x31.9=3.7

QUALITY CONTROL IN THE MANUFACTURE OF
MINIATURE PRECISION BEARINGS

Charles J. Hudson
Miniature Precision Bearings, Inc.

The bearings manufactured by Miniature Precision Bearings, Inc. are classed as a high precision product. Requirements by customers are exacting. Dimensions must be held to close tolerances. Performance characteristics must be held within close limits which are unusual and interesting to people unfamiliar with the precision ball bearing industry.

The bearings about to be described are known as ball bearings. They consist of three major parts (1) an outer race, (2) an inner race and (3) a specified quantity of balls between the two. Figure 1 shows a full race ball bearing. Many ball bearings have retainers or separators between the balls. As the bearing rotates in use the balls rotate in ball grooves ground in both races.

In use either the outer or inner component is usually stationary while the remaining component and balls rotate. In some adaptations loads are relatively high and speeds of rotation may be up to 60,000 R.P.M. Long life, low power consumption and accurate movement are the chief requirements. In other uses, in delicate instruments the rotation of either component may be very nearly zero with the chief requirements being extreme accuracy and low starting torque. In most cases these bearings are used where miniaturization of equipment and light weight are prime requirements.

The largest bearing manufactured by Miniature Precision Bearings, Inc. is 3/8" outside diameter. The smallest made at present is .100" outside diameter complete with inner and outer races plus very small balls. In the very near future even smaller ones will be available whose outside diameter will be .0625". It will take several hundred of these very small bearings completely assembled to fill an ordinary thimble.

Aside from the full race radial type bearing illustrated in Figure 1 there are many other types such as thrust, pivot and angular contact bearings. Variations in type together with dimensions, many with special requirements, result in several hundred different bearings being on the available manufacturing list.

These precision bearings are particularly well adapted for use in precision instruments, in delicate mechanisms where light weight, long life bearings of miniature size are required. Some of the adaptations are in drive movements for recorders, time clocks, barometers, meters, gyroscopes and instruments of delicate responsiveness. They have high load capacity so can be used where jewel pivots would fail because of low physical strength.

Although standard machine tools are used in manufacturing these miniature bearings wherever possible, their small size and close accuracy means that many of the tools used in manufacture and inspection must be made within the plant. A good tool room and an ingenious tool maker are necessary for economical manufacture. Gages for measuring physical characteristics of ball bearings of usual sizes are available

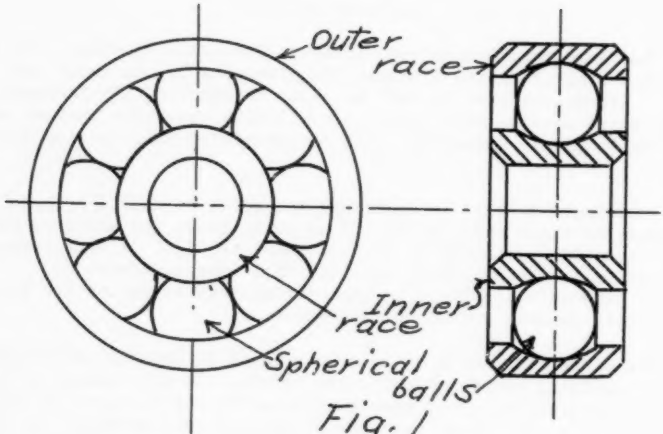

Outer
race
Inner
race
Spherical
balls

Fig. 1

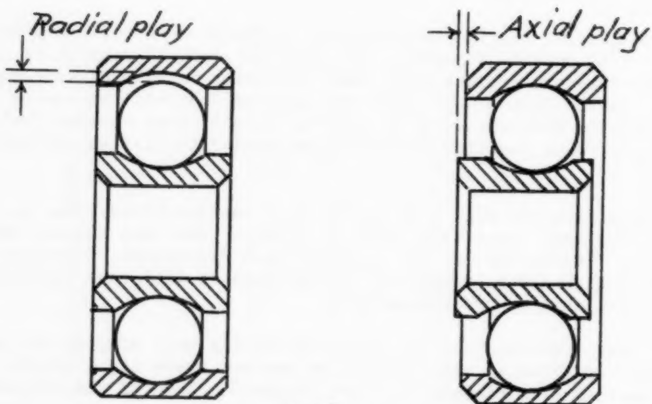

Radial play
Axial play

Fig. 2

on the market. However many of these gages are not adapted for use in the manufacture of miniature bearings or are very special and prohibitive in cost.

Because of the very small sizes of these bearings the job of handling them while manufacturing, measuring and inspecting them is serious. Much of the work has to be done with the aid of microscopes and with special equipment adapted to use with these miniature bearings.

Quality control studies are sometimes said to be impracticable in a small plant, in a plant where high precision is required and where continuous runs of the same product are unavailable. Studies at Miniature Precision Bearings, Inc. refute these arguments. The plant is small, employing less than 200 people. Precision requirements are high with most dimensions being maintained to within .0002" and some to within .0001". Measuring equipment must be maintained to within .000020" of nominal size. About 200 different types of bearings are being manufactured and are considered standard. Many bearings of special dimensions and shapes are also made in addition. Long runs of large quantities are scarce.

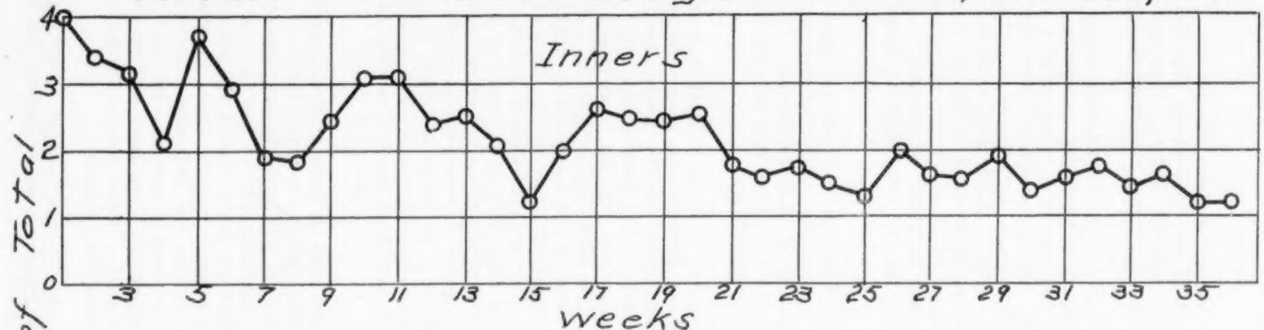
It has been demonstrated that Quality Control methods are applicable to the miniature ball bearing industry. It is believed that better customer-vendor understanding exists because of Quality Control activities and that satisfactory government relations are present where orders are accompanied with surveillance requirements.

During the period of Quality Control activity at Miniature Precision Bearings, Inc. the amount of rejections and reworks has been reduced by about fifty percent. This has been accomplished by the use of Quality Control methods involving various kinds of control charts, records, tables, etc. In addition, however, much credit for the improvement must be given to operators and supervision. The fact that a Quality Control program was in existence has tended to make all personnel more quality conscious, exercise more care in manufacture with the result that all product is made at a higher quality level as well as with a lowering of rejections and reworks. It is probably a fact that among smaller groups of employees which exist in this kind of a plant it is a necessity that everyone become his own inspector to a greater degree than in large plants. Quality Control education promotes a critical attitude in operators and encourages pride in workmanship. Figure 3 shows the reduction in rejections and reworks over a period of several months since Quality Control methods were started.

At Miniature Precision Bearings, Inc. use is made of frequency distribution charts, of weekly and monthly defective records and of capability studies. At times the standard $\overline{X}$ and R charts are used to study machine capability. This is done particularly at the time of arrival of new machines. Examples of some of these charts and records follow. Management and manufacturing personnel receive these charts and records. Monthly meetings are held to discuss the evidence. Records are obtained by roving inspectors taken at hourly periods, by sampling inspection of component parts in the Inspection Department and by final inspection also in the Inspection Department. Basic cost information in dollars of rejections and reworks complete the picture.

The first recorded step in the Quality Control effort is at the automatic screw machines where both the inner and outer components are

Miniature Precision Bearings, Inc. Keene, New Hampshire



Graphs show decrease in rejected components during a period of eight months of Quality Control activity

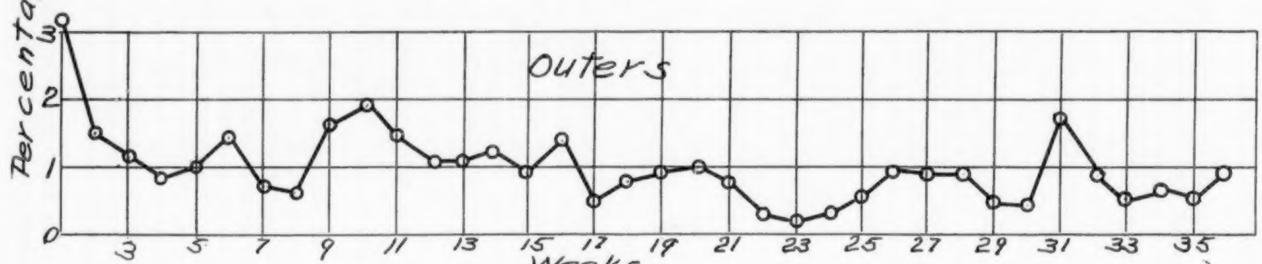


Fig 3

SIZE & TYPE Inner 55518 LOT NO: 488

CHAMFER	FILLER SLOT	MUTILATION	GROOVE FIN.	CORROSION	GEN. APPEAR.	TAPER	BORE O.S.	O.D.	CONCENT.	OUT OF ROUND	WIDTH	MISC.	CRACKED	BELLMOUTH	DEF. BORE	FACE DEF.		RECHAMFER	FACE REP.	LAND REP.	O.D. CHAMFER	BORE U.S.	REJECT	REPAIR	ACCEPT		TOTAL
3						3	1									2		6					9	6	753		768
						1				1													2		107		109
23						4	1	1		3					1	10		15	17	1		1	43	34	1490		1567
						2	2								5	3		10	6				12	16	155		183
						5	2								2	13		38	46				22	84	838		944
2						6										7		3	7			1	15	11	513		539
						1				2													3		111		114
							2												4				2	4	598		604
						1	1									3							5		601		606
							1			2													3		467		470
28						23	10	1		8					8	38		72	80	1		2	116	155	5533		5804

Fig. 4

machined from solid bar stock. At this point a roving inspector measures a sample of ten pieces each hour. A simple p-chart of findings is kept for each machine. The several characteristics measured on each consist of inside and outside diameters, width, diameter and location of ball groove and general appearance. Percent defectiveness, if any, is recorded on the p-chart where it can be observed by operators and supervision. Excessive discrepancies are immediately brought to the attention of supervision.

After hardening, and lapping the sides, the subsequent operations are grinding and honing to bring dimensions to within final requirements before assembling. Here again roving inspectors make hourly inspections for final dimensions and general appearance of component parts. These inspection results are recorded on p-charts which in turn keep operators and supervision informed as to whether or not limits are being met.

Figure 4 shows a monthly running record of performance of one of the component parts. It is from such records as these that causes of defectiveness are located and corrections made so that the number of rejections and reworks can be reduced. This chart is presented to illustrate the fact that Quality Control methods can be applied to precision manufacture of miniature ball bearings just as in other types of industry. It can readily be observed that at the time this record was made three operations were the cause of major rejection and repair work, taper of the bore, rechamfer and repair of faces.

Among several requirements of performance of the finished ball bearings are concentricity of both inners and outers, radial play of the outer with respect to the inner, starting torque and running torque. These are all measurements of performance in the intended use of the bearings. (See Figure 2)

There are standard requirements for some of these properties set up by The Anti-Friction Bearing Manufacturers Association, Inc. for various classes of bearings. Those manufactured by Miniature Precision Bearings, Inc. conform to the requirements of Class 5 or 7 which are of the highest precision. In addition, some customers ask for even closer precision requirements.

Figure 5 is a histogram representative of a test for concentricity of standard assembled bearings. Manufacturing requirements specify that concentricity should be within .0002". The chart shows the majority of these bearings to run true to within two tenths of a thousandth of an inch.

Figure 6 is a representative histogram showing the results of a test for radial play. Miniature Precision Bearings, Inc. requirements for Class 5 bearings specify that radial play should not be greater than .001" and must be greater than zero. It frequently happens, however, that some customers ask for a particular range of radial play within rather narrow limits, as for example from .0001" to .0003"; others might ask for a range of from .0002" to .0004", etc. In Figure 6 it will be noted that a high percentage of the bearings show a concentricity between .0001" and .0003". Those outside the range of .0001" to .0003" are not rejections. They can be used to fill other requirements as long as none have a radial play greater than .001".

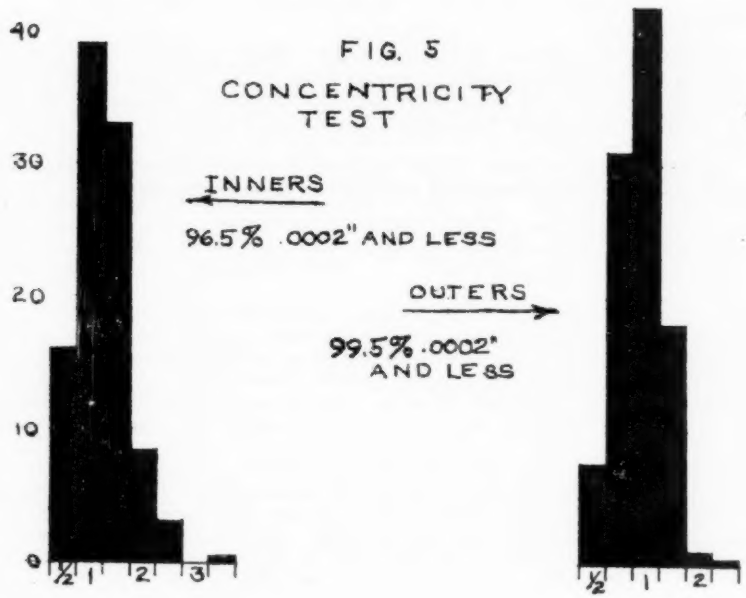


FIG. 5
CONCENTRICITY TEST
INNERS
96.5% .0002" AND LESS
OUTERS
99.5% .0002"
AND LESS
40
30
20
10
0
½ 1 2 3
½ 1 2

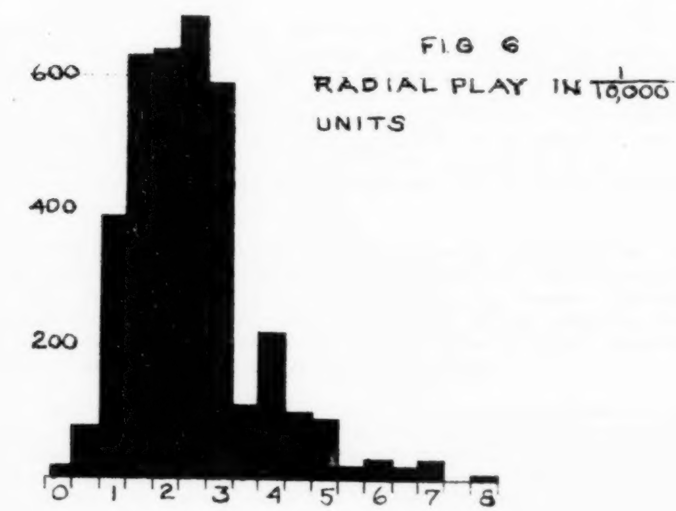


FIG 6
RADIAL PLAY IN 1/10,000
UNITS
600
400
200
0 1 2 3 4 5 6 7 8

Much of the product manufactured by Miniature Precision Bearings, Inc. finally enters into government use, and is subject to government surveillance. Sampling plans in use are the usual MIL-STD-105A.

During the past few months a punched card system has been installed which is destined to be of great help in furthering Quality Control efforts.

The above description is rather brief. There is still more work to be done to complete the program. However the value of Quality Control efforts is recognized by management and manufacturing personnel. To force efforts too rapidly can result in slow progress.

Acknowledgment is made to the management of Miniature Precision Bearings, Inc. for permission to publish the above experiences.

CORRELATION ANALYSIS IN BATCH
PROCESS CONTROL

John D. Hinchen
Monsanto Chemical Company

In order for a finished product to have consistent high quality, the desirable characteristics have to be "built in" all along the line. Statistical Quality Control exists, and is successful, mainly because it assists the manufacturer in performing the very vital task of making the quality right the first time through. However, much of the literature on the use of statistical methods in product quality control has been devoted to development of the techniques for separating the two basic types of variation - that which is inherent to the process, and that which is indicative of new or changing causes of quality variation. The main advantage of these techniques is that they tell the production man when an "assignable cause" of variation exists in the process. The task of locating and correcting this cause is not necessarily assisted by the ordinary quality control methods.

In many chemical processes, this type of control is usually an "after the fact" proposition - the basic control of quality must be handled while the product is being manufactured. The people responsible for the control of finished product quality find that their chief responsibility lies in guiding the course of the reaction so that the desired end goal is reached in each batch of material.

Many statistical techniques are available and can be of use to the manufacturer for establishing the required control of operating conditions, and pinpointing the critical turning points in the reaction. One of the most valuable of these methods is correlation analysis, and we hope to show by example how this useful statistical technique can help to establish more critical control of process performance.

In the manufacture of a certain type of synthetic liquid resin, some of the properties of importance are the mineral spirits tolerance and the solids content. For some end uses, too little mineral spirits tolerance precludes the use of the material in the optimum concentrations required by the customer; too great tolerance indicates a low molecular weight product which may have processing characteristics that are in many ways undesirable. Solids content control is important to both the manufacturer and the user. A uniform resin concentration helps both of these people maintain predictable cost control, and provides uniform handling properties.

The reaction process for the resin under consideration consists of two phases. In the first phase, the monomer is reacted to a certain stage. This controls, among other things, the mineral spirits tolerance of the finished product. The second stage of the reaction consists of the removal of the water present. The amount of water removed will control the solids content of the end product.

The process as originally developed was established to produce a resin which varied somewhat widely in dilution characteristics. As we became more skilled in evaluating the material, we began to find that certain batches of resin enabled us to use optimum operating conditions, resulting in a better yield and reduced operating cost. Working together with the end user, it was determined that resin having

UNTRANSFORMED RELATION

BETWEEN MINERAL SPIRITS TOLERANCE

AND UNREACTED MONOMER AT CHANGEOVER

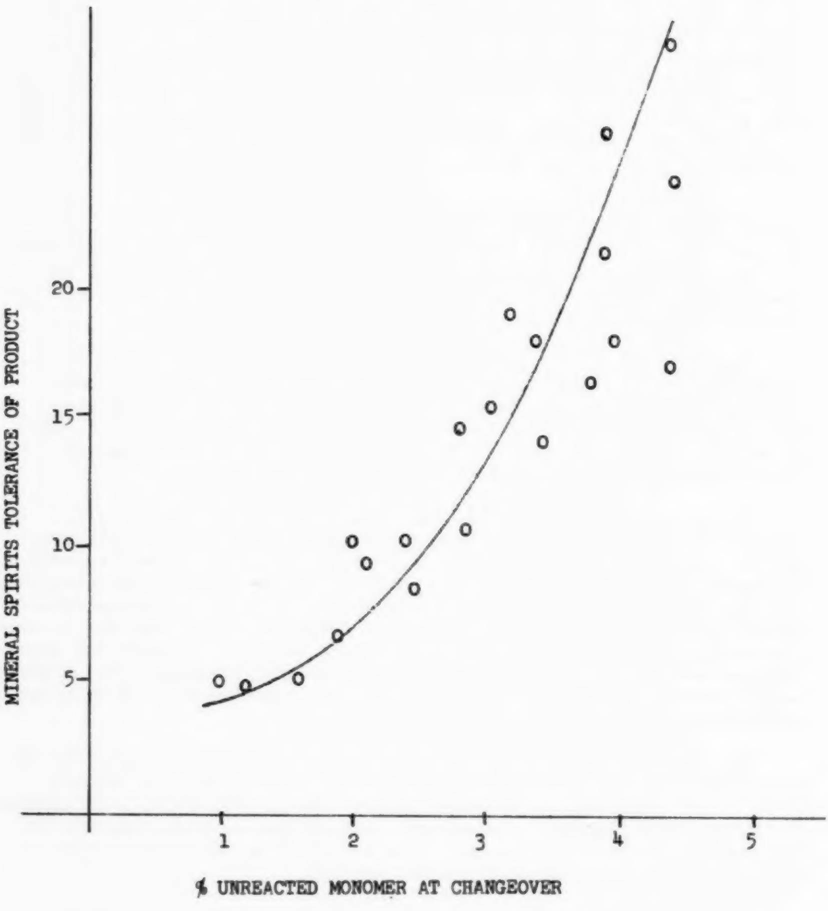


FIG. 1

LINE FOR ESTIMATING MINERAL SPIRITS TOLERANCE FROM

UNREACTED MONOMER AT CHANGEOVER

(LOG TRANSFORMATION USED)

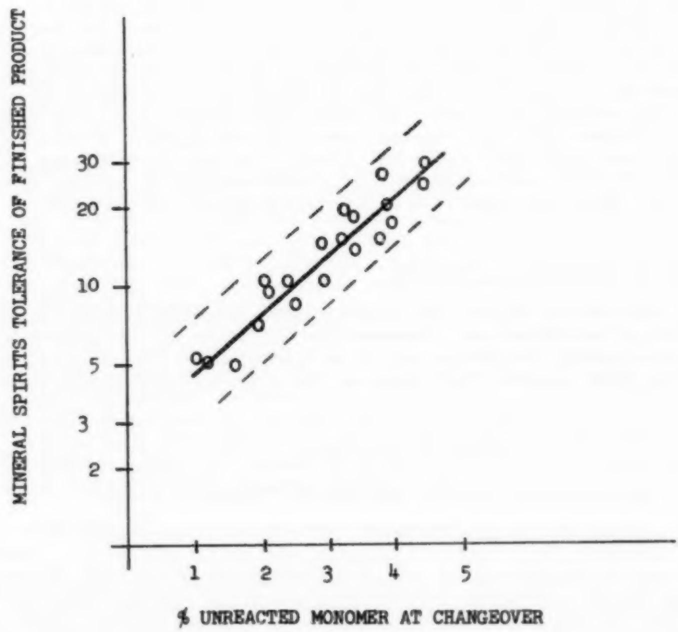


FIG. 2

tolerance values of between 10 and 20 was the most desirable. These limits were established as tentative manufacturing specifications. It became the problem then to control the initial portion of the cycle very closely, so that the finished product would fall within the required limits.

The progress of the initial portion of the cycle was followed by testing the amount of monomer remaining unreacted in the reaction mixture. The amount of unreacted monomer became gradually less as the reaction proceeded, and with reagent and catalyst concentrations fairly well controlled, had a fairly good relationship with the average molecular weight of the finished product. In order to determine the control point at which the reaction would be quenched, it was necessary to study process history to determine the monomer content at changeover (end of reaction cycle) for those batches of resin which had the desired final properties.

Table I shows the monomer content at changeover, and the final tolerance of twenty lots of resin. By plotting these values on a scatter diagram it was found that the relationship between the two properties was curvilinear (See Figure 1). Since the tolerance was expressed as a ratio (volumes of naphtha to volumes of resin to produce precipitation), it was felt that some advantage would be gained by plotting the variable on a ratio or logarithmic scale. Figure 2 shows that the relationship thus obtained appears to be linear. To verify this conclusion, and to establish the relationship between free monomer at changeover and final tolerance, the data were analyzed by simple linear correlation, using the logarithm of the tolerance as the dependent variable.

Discussion of Correlation Analysis

The correlation coefficient as calculated from the data below was 0.96, showing the relationship between the logarithm of the tolerance, and the free monomer at changeover to be significant at 99.9% confidence level. The least squares line representing the relationship was found to be

$$y = 0.45 + 0.22x$$

where y = log tolerance (volumes naphtha per volume of resin)

x = monomer at changeover (%).

The 95% confidence limits around the correlation line were $\pm$ 0.14 on the log scale, and the line and limits are shown together with the plotted data on Figure 2. The data from which the statistical conclusions were drawn follows:

TABLE I

% Monomer at Changeover	Mineral Spirits Tolerance	Log. M. S. T.
1.0	5.0	0.70
1.2	4.9	0.69
1.6	5.0	0.70
1.9	7.0	0.84
2.0	10.2	1.01

TABLE I (Cont'd.)

% Monomer at Changeover	Mineral Spirits Tolerance	Log. M. S. T.
2.1	9.5	0.98
2.4	10.5	1.02
2.5	8.5	0.93
2.9	11.0	1.04
2.8	14.5	1.16
3.1	15.0	1.18
3.2	19.0	1.28
3.4	14.0	1.15
3.4	18.2	1.26
3.8	16.3	1.21
3.9	18.0	1.25
3.9	21.0	1.32
3.9	26.0	1.41
4.4	29.0	1.46
4.4	24.0	1.38

From the results of the foregoing correlation analysis, the optimum changeover point to produce the desired M.S.T. was 3.3% free monomer. The next step was to make the required changes in standard procedure to enable the production department to use the new control point.

Due to the residual error around the regression line, it was necessary to make the changeover during a very narrow range of free monomer in order to obtain the greatest percentage of material within the required specifications. Up to this time, the procedure had been to plot the monomer content versus the reaction time, and to predict from the slope of the line, the probable time at which the desired changeover point would be reached. Figure 3 shows the plotted values for twenty batches of material. To avoid complicating the diagram, the lines joining the points for each batch are omitted. Although there are some differences in the slopes of the lines, the greatest difference is in the levels of monomer contents at a given reaction time. These differences are caused by partial reaction taking place during the time the batches were being heated up to reaction temperature. The differences in slope are caused by normal variations in temperature, catalyst concentration, monomer purity, etc.

This method of predicting the end-point was not entirely satisfactory. To obtain acceptable accuracy of prediction, a fairly large number of samples had to be obtained, and the dangers of extrapolation were always present. It was the production supervisor who pointed out the answer to the problem. He reasoned that we were not particularly interested in the history of the reaction up to the point of test, but primarily in how much further the reaction still had to proceed before the desired monomer concentration was reached. In other words, the variable to be predicted was not free monomer at a given reaction time, but the amount of time remaining in the reaction period, to be predicted from the measured free monomer content. The correlation analysis should be made between free monomer and "minutes to changeover", as this variable was named. However, as pointed out by Windsor ("Which Regression?" Biometrics Bulletin, Vol. 2, 1946), the time factor is still the independent variable, even though it is the variable which is being predicted. We must use the line for predicting monomer content from time, in order to obtain our best estimate of the time remaining in the

FREE MONOMER CONTENT

VS.

REACTION TIME

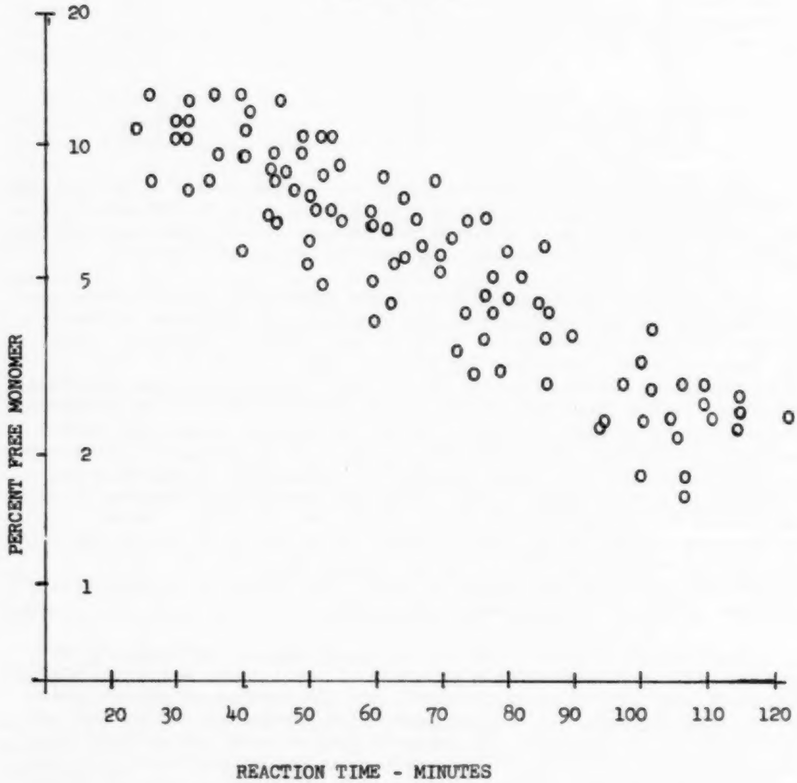
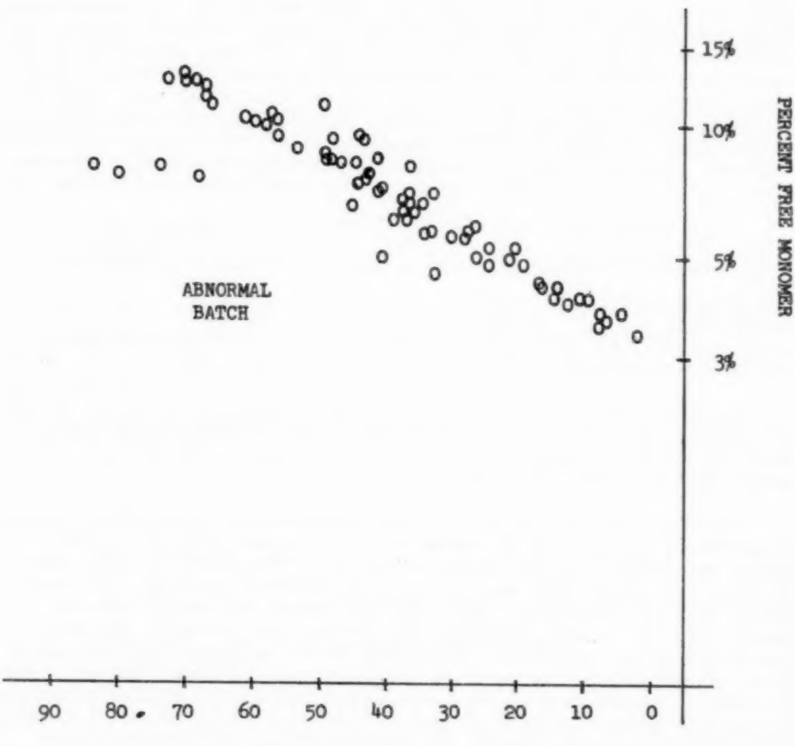


FIG. 3

FREE MONOMER CONTENT

VS.

MINUTES TO CHANGEOVER



MINUTES TO CHANGEOVER

FIG. 4

reaction. Using the same data as in Figure 3, but substituting minutes to changeover for reaction time, the relationship shown on the scatter diagram in Figure 4 was obtained.

In analyzing the available data from these twenty batches of resin, it was found that the relationship between free monomer and reaction time was not linear, but followed a logarithmic curve. The rate of polymerization was evidently proportional to the amount of the monomer remaining free in the reaction mixture. Before running the correlation analysis, the monomer values were converted to logarithms, making the principles of simple linear correlation applicable.

Discussion on Correlations

The correlation analyses were run using all of the available data, and both relationships were found to be highly significant. The residual error around the correlation line, was much smaller for the relationship which used minutes to changeover as the X variable in place of reaction time. When these data were plotted, it was found that one batch of material gave results which were consistently outside the 95% confidence limits around the line. On investigating the history of this batch it was found that a non-normal situation involving poor heat transfer existed, and that the rate of reaction was much slower than would normally be expected. The data from this batch were set aside and the correlations recalculated using the remaining data. The results were as follows:

	Log Monomer Vs. Reaction Time	Log Monomer Vs. Minutes to Changeover
Correlation coefficient	0.80	0.98
Significance	99.9%	99.9%
Least Squares line	$y = 1.263 - 0.0073X_1$	$y = 0.518 + 0.0086X_2$
95% limits of prediction	$\pm$ 0.198	$\pm$ 0.066

Figures 3 and 4 show the two relationships plotted on semi-logarithmic graph paper. The scale is reversed on the "minutes to changeover" graph to facilitate comparison, and to make usage of the chart in the normal left to right direction.

The much greater precision of the relationship between free monomer (logarithms) and minutes to changeover, led to the adoption of this method for predicting the changeover point. Forms based on semi-logarithmic graph paper were made up with the standard reaction line, and made a part of the process logs. As each sample was taken from the kettle, and the results obtained and plotted, the probable remaining reaction time was estimated. Depending on the length of this estimated time, the decision as to sampling frequency was obtained. It was possible after a period of time, to reduce considerably the number of samples required to obtain accurate changeover control. The procedure has since been extended to other resins, and has materially improved the right first time record.

The customer is also vitally interested in accurate control of the solids content of the finished product. This property is controlled during the second portion of the cycle, in which water is removed from the kettle by vacuum dehydration. When the material was originally formulated, the tentative process included estimation of solids content during the course of the reaction by running a specific gravity measure-

ment on the resin solution. Figure 6 shows the relationship that existed between the solids content and the specific gravity of several finished batches of resin.

It was decided that the relationship between these two properties was not sufficiently precise (although highly significant), to enable estimation of solids content with the required degree of accuracy. Normal practice was to chart the progress of the reaction by specific gravity until the estimated solids content had been reached. Then the dehydration was stopped, and a sample taken for a 3 hour solids test. If the result was within specifications the resin was packed out; if not, further dehydration or water addition was made for solids adjustment.

In many cases, the specific gravity was affected by factors other than solids, and waiting for the solids test for 3 hours was often justified. However, this was costing money and processing time, so a new test for predicting solids more accurately was needed. In addition, a more precise test for solids content was developed and used as a standard test for this product. The lab came up with the suggestion that refractive index be tried, and Figure 5 shows the resultant relationship.

When the relationships were plotted on square-ruled graph paper it was evident that linear correlation techniques would apply without transformation. The correlation coefficients were calculated, and the following results were obtained:

1. For the solids vs. specific gravity analysis, $r = 0.75$

2. For the solids vs. refractive index analysis, $r = 0.94$.

The first relationship explains 56% of the total variance of the solids content, leaving a residual variance of 44% unexplained. The second relationship is considerably better, leaving only 11% as unexplained residual variance.

The two correlation lines were as follows:

1. For the solids vs. specific gravity analysis.

   % solids = 421 (specific gravity) - 475.5

   95% limits $\pm$ 2.7%

2. For the solids vs. refractive index analysis

   % solids = 305.6 (refractive index) - 345.4

   95% limits $\pm$ 1.3%

The data used for these calculations are plotted on the attached illustrations.

Due to the increased confidence in the predictability of the solids content, two things were accomplished:

1. It was possible to pinpoint the end of the dehydration cycle more accurately, and produce a material with a more uniform solids content.

LINE FOR ESTIMATING SOLIDS

FROM REFRACTIVE INDEX

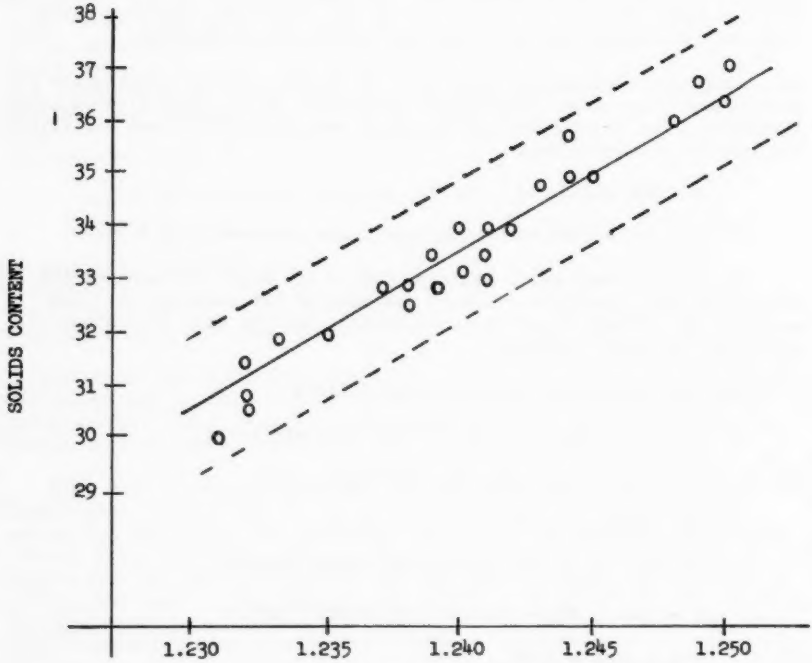


FIG. 5

LINE FOR ESTIMATING SOLIDS CONTENT

FROM SPECIFIC GRAVITY

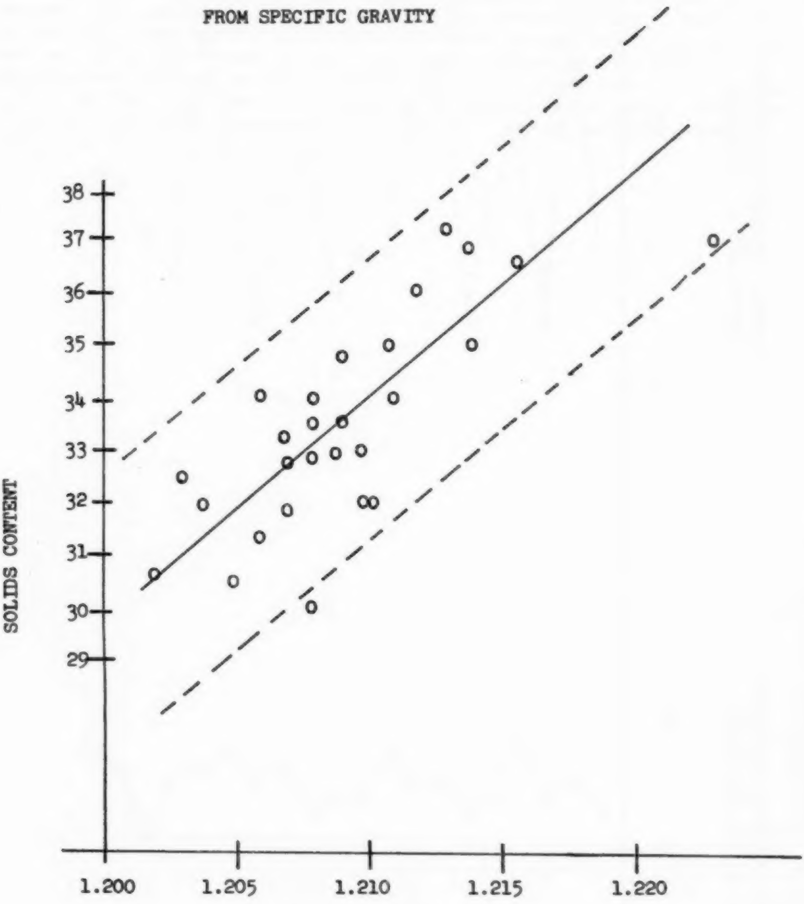


FIG. 6

EFFECT OF MINUTES TO CHANGEOVER CONTROL

ON MINERAL SPIRITS TOLERANCE

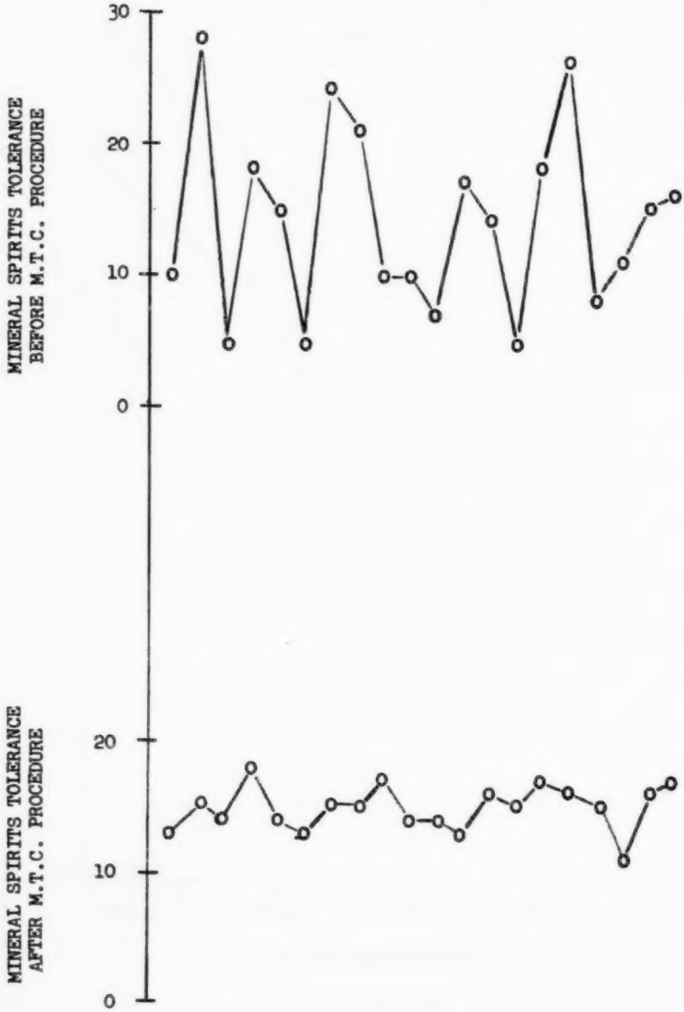


FIG. 7

2. The greater accuracy made it possible to drain the batch on the basis of refractive index alone, without waiting for the time-consuming solids test.

By the use of the correlation techniques, better quality and lower costs were achieved.

Finally, the test data on the finished product after the new procedures were installed, were compared to the previous product level and variability. The control charts for dilutability and solids content before and after the change showed clearly the effect of this approach on the finished product quality. Further application was made to other products, and the department was placed on a sounder basis quality-and-cost-wise, with the added advantage of increased customer confidence and satisfaction.

BINOMIAL PROBABILITY TABLES

Calvin J. Kirchen
Lincoln-Mercury Division
Ford Motor Company

Several well-known texts on statistical quality control either directly, or by implication, attach the word burdensome to direct application of the binomial distribution to sampling inspection problems. This distribution, represented analytically by the expression $(Q+P)^n$, applies to the drawing of a sample of n pieces from a lot with fraction defective P, said lot being so large that the drawing of the successive items does not affect P (Q, of course, is the fraction of the lot which is good, so $Q=1-P$). The constancy of P theoretically requires a lot with an infinite number of pieces in it, but in practice sufficiently accurate results are obtained if the lot size is at least ten times the sample size.

Until recently the Poisson distribution has been "pulled from under the counter" as a substitute for the well-known, but rarely applied, binomial distribution. This is no longer necessary now that two sets of tables of the latter distribution are available to the general public:

1. National Bureau of Standards: "Tables of the Binomial Probability Distribution"—NBS AMS6, Government Printing Office, Washington 25, D. C., 1949.

2. Romig, Harry G.: "50-100 Binomial Tables", John Wiley and Sons, Inc., New York, N. Y., 1952.

Since the NBS Tables cover sample sizes by unit intervals from 2 to 49 and Romig's "50-100 Binomial Tables" cover sample sizes from 50-100, in increments of 5, these two tables do not overlap with respect to sample size, they provide a very valuable source of theoretical values of this important distribution. Fig. 1 on the following page is an excerpt from the NBS Tables to indicate their format. This format suggested to me the possibility of preparing tables with fewer decimal places, with sample sizes in more common usage in quality control applications, and with a somewhat different arrangement of columns. Fig. 2 on the second following page gives the table for sample size=n=10 which I prepared as an example of the complete set. Copies of the complete set are available to those at this session, or they may be obtained by writing me at my business address.

The tables I prepared, with numerical values derived entirely from the NBS Tables, have the following characteristics:

1. Sample size, n: 4, 5 and all multiples of 5 through 45.

2. Lot fraction defective, P: .01 to .10 by .01; by .05 from .10 to .30; by .10 from .30 to .50.

3. Probabilities of the occurrence of exact numbers of defectives in the various sample sizes, given in columns headed by various values of x.

4. Probabilities of the occurrence of at most a specified

FIG. 1

p	n=4 r=2	n=4 r=1	n=4 r=0	n=5 r=4	n=5 r=3	n=5 r=2	n=5 r=1	n=5 r=0
.01	.0005880	.0388120	.9605960	.0000000	.0000099	.0009702	.0480299	.9509900
.02	.0023050	.0752953	.9223682	.0000008	.0000768	.0037648	.0922368	.9039208
.03	.0050808	.1095208	.8852928	.0000040	.0002540	.0082141	.1327939	.8587340
.04	.0088474	.1415577	.8493466	.0000123	.0005898	.0141558	.1698693	.8153727
.05	.0135376	.1714750	.8145062	.0000297	.0011281	.0214344	.2036266	.7737809
...								
...								
...								
.50	.3750000	.2500000	.0625000	.1562500	.3125000	.3125000	.1562500	.0312500

FIG. 2

	P	x=0	x=1	c=1	x=2	c=2	x=3	c=3	x=4	c=4	x=5	c=5	x=6	c=6	x=7	c=7
n=10	.01	904	091	996	004	1000										
	.02	817	167	984	015	999	001	1000								
	.03	737	228	965	032	997	003	1000								
	.04	665	277	942	052	994	006	1000								
	.05	599	315	914	075	988	010	999	001	1000						
	.06	539	344	882	099	981	017	998	002	1000						
	.07	484	364	848	123	972	025	996	003	1000						
	.08	434	378	812	148	960	034	994	005	999	001	1000				
	.09	389	385	775	171	946	045	991	008	999	001	1000				
	.10	349	387	736	194	930	057	987	011	998	001	1000				
	.15	197	347	544	276	820	130	950	040	990	008	999	001	1000		
	.20	107	268	376	302	678	201	879	088	967	026	994	006	999	001	1000
	.25	056	188	244	282	526	250	776	146	922	058	980	016	996	003	1000
	.30	028	121	149	233	383	267	650	200	850	103	953	037	989	009	998
	.40	006	040	046	121	167	215	382	251	633	201	834	111	945	042	988
	.50	001	010	011	044	055	117	172	205	377	246	623	205	828	117	945

P	x=8	c=8	x=9	c=9	x=10	c=10
.30	001	1000				
.40	011	998	002	1000		
.50	043	989	010	999	001	1000

number of defectives in the various sample sizes, given in columns headed by various values of c.

Although my tables have their numerical roots in the NBS Tables, it is with pleasure that I acknowledge my debt to Dr. Romig's "50-100 Binomial Tables" for the idea it gave me to make up tables for the multiples of 5 less than 50, omitting most of the other sample sizes available in the NBS Tables. As a matter of fact, the completeness of the NBS Tables contributes to making them a reference volume rather than a handy tabulation for the quality control practitioner. Values for sample size 4 were included because of its frequent use in industrial quality control.

HOW TO READ THE TABLES

Reading of the tables will be illustrated with Fig. 2, which, of course, is confined to sample size, n, = 10. The tabular values refer to individual values and sums of them in the following binomial expansion:

$$(Q+P)^{10} = Q^{10} + 10Q^9P^1 + 45Q^8P^2 + 120Q^7P^3 + 210Q^6P^4 + 252Q^5P^5$$
$$+ 210Q^4P^6 + 120Q^3P^7 + 45Q^2P^8 + 10Q^1P^9 + P^{10}$$

Any college algebra book will provide information on how to form individual terms for other values of n.

Values of P are given in the first column; Q=1.00-P or is the fraction of the lot which is acceptable. The exponent of the binomial (Q+P) is, of course, the sample size. Assuming a specific value for P of, say, .04, the value .665 in the row for .04 and in the column headed x=0 is the value of $Q^{10} = (1.00-.04)^{10} = .96^{10}$. The interpretation is: the probability is .665 that a sample of 10 drawn at random from a lot of a least 100 pieces will have no defective pieces in the sample; that is, will have only good pieces. In the next column, headed by x=1, the tabular value of .277 is interpreted as the probability of the occurrence of exactly 1 defective piece among the sample of 10. The value .942 in the column c=1 is the probability of the occurrence of 0 defectives or 1 defective and hence is .665 + .277=.942. All tabular values have a maximum error of .0005, and blank spaces in rows after the apperance of 1.000 mean that the probability of occurrence is less than .0005. It is well to note that the exponent of P in any term is identical with the value of x in the column of the tables which gives the value of that term.

The tables for n=4 and n=5 reveal a surprising fact: whether one takes a sample of 4 or 5 pieces from a lot which is .20 defective (P=.20), the probability that the sample will have one defective piece in it is the same, namely, .410! That this is so can be proved by a comparison of the terms in the two binomial expansions which give probabilities of exactly 1 defective in each size sample:

$$n=4:\ 4(.80)^3(.20)^1 = .4096, \text{ or } .410$$
$$n=5:\ 5(.80)^4(.20)^1 = .4096, \text{ or } .410$$

While values of P apparently have been restricted to a maximum of .50, it is possible to "read" the tables for values of P greater than .50: fractions defective which are complementary to the given ones of the

table; that is, the two add up to 1.00. For example, P=.60 can be used, as .60+.40=1.00, and .40 is an original value of P; but P=.65 can't be used because its complement, .35, is not an original value of P in these three decimal tables. When we "read" the tables for P greater than .50, the roles of Q and P with regard to their numerical magnitudes are reversed, and this has the effect of changing the meaning of the column headings from referring to the number of defective pieces occurring in the specified sample to the number of acceptable pieces in the sample. For a specific example of this: take P=.60; find the row corresponding to .40, which is now the fraction of the lot which is acceptable. For n=10, Fig. 2, .006 in the x=0 column is the probability of the occurrence of 0 acceptable pieces, or 10 defectives; .040 in the x=1 column is the probability of the occurrence of exactly 1 acceptable piece, or 9 defective pieces; and .046 in the c=1 column is the probability of occurrence of at most 1 acceptable piece, or at least 9 defective pieces. Note the shift from "at most so many acceptable pieces" to "at least so many defective pieces".

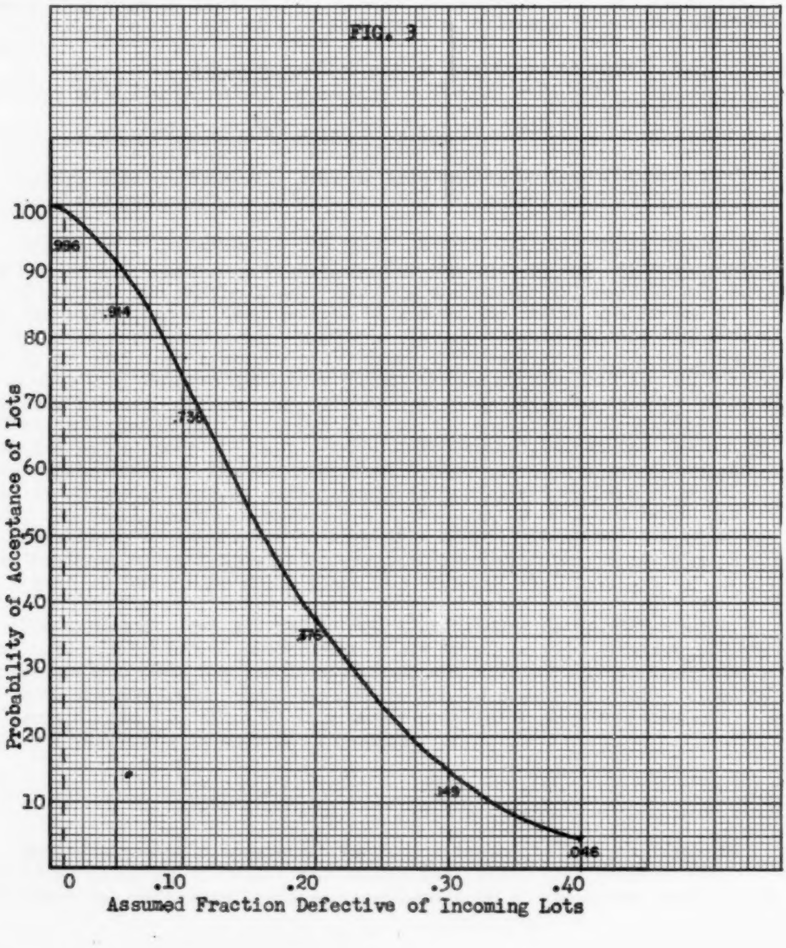
(Paper is continued on the following page)

USES OF THE TABLES

Besides the use of the tables given in the preceding material, there are "column-wise" and row-wise" uses of the tables, which will be described and illustrated in this section:

1. Column-wise use--construction of Operating Characteristic Curves of single sampling plans. Plot probabilities obtained from columns headed by x=0 or any value of c, depending upon the acceptance number of the sampling plan, against values of P. For example, if we take n=10 and permit one defective in the sample for acceptance, we plot the probabilities in the column <u>c=1</u> on a vertical scale against values of P on a horizontal scale. See Fig. 3.

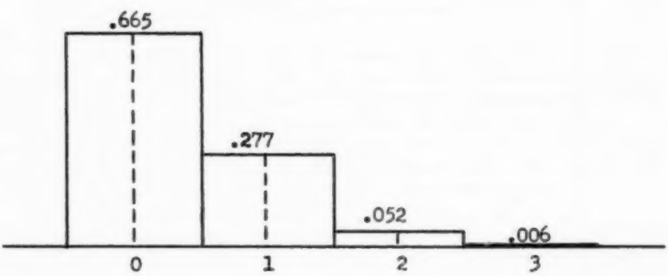
FIG. 3



2. Row-wise use—construction of histograms of specific binomial distributions. For a given value of P, probabilities corresponding to the occurrence of exact numbers of defectives in a sample of size n are made the altitudes of rectangles whose bases all are of unit length; the midpoint of each rectangle is marked to correspond to the exact number of defectives for which the altitude is the probability of occurrence. As an example, take n=10, P=.04, and form the following table:

Base midpts.	Altitude
x=0	.665
x=1	.277
x=2	.052
x=3	.006

FIG. 4

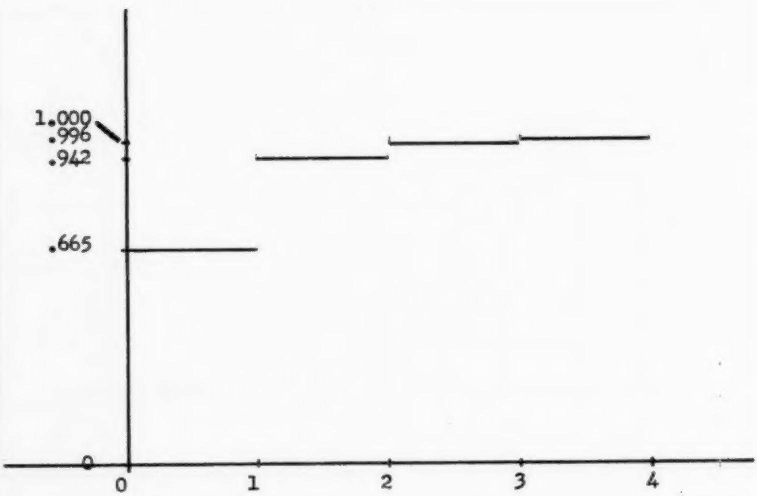


Another "row-wise" use of the tables is to plot the ogive, or cumulative frequency function, or distribution function for a particular binomial distribution. The tabular entries in the columns of cumulative values, including x-0, are plotted against the observable number of defective pieces, as is shown in the figure below, which is based on the following table:

n=10, P=.04

Obs'vd No. of Defectives	Cumulative Frequency
0	.665
1	.942
2	.994
3	1.000

FIG. 5



WHERE THE BINOMIAL TABLES CAME FROM

I am sure it will be of interest to know that the NBS Tables were obtained from "Tables of the Incomplete Beta Function", edited by K. Pearson, The Biometrika Office, University College, London, 1934. The NBS version of these tables is contained in Table II of that volume. Table I of that volume, of which Fig. 1 of this paper is an excerpt, was obtained by taking differences between successive entries in Table II. While the probabilities in columns of my tables corresponding to individual terms in the binomial expansions are three-decimal forms of the seven-decimal values of NBS Table I, the cumulative values are not similarly related to NBS Tables II. Rather, my cumulative values are complements of corresponding ones in Table II.

Pearson's"Tables of the Incomplete Beta Function" have been used in other ways than for the NBS Tables:

1. Gen. Simon based his IQcharts which appeared in his "An Engineers' Manual of Statistical Methods", John Wiley and Sons, Inc., N.Y., N.Y., 1941, on them;

2. Dr. Frank E. Grubbs based tables in his paper "On Designing Single Sampling Inspection Plans," Annals of Mathematical Statistics, 20 (1949), 242-56, on them, the tables giving sample sizes, acceptance numbers and values of the lot fraction defective for which probabilities of acceptance are .95 and .10.

3. Robert E. Clark, The Pennsylvania State College, used them in his paper "Percentage Points of the Incomplete Beta Function", Journal of the American Statistical Association, 48(1953)831-43.

Gen. Simon identified his set of charts by means of a probability level for each one: .995, .90, .50, .10 and .005. Binomial probabilities and the several charts are "interchangeable" only when a tabular probability is the complement of one of Gen. Simon's five probability levels. For example, for n=10, P=.01, Prob (x=0) =.904, whose complement is approximately .10; then on Simon's Chart 0.1, the curve c=0 intersects the ordinate at n=10 at fraction defective=.01.

In Dr. Grubbs' Table can be determined, for a given sample size and allowable number of defective pieces, that lot fraction defective for which the probability of acceptance is .95, or .10, as the case may be. Entering his .95 table with n=10 and c=3, we read .15 in the body of the table; in the binomial tables, we enter them for n=10, c=3, and in the row for P=.15, we read .950 for the corresponding probability. Dr. Grubbs' Tables and my binomial tables exchange roles of P and of the correspodning probability. Interchangeability is at a minimum because Dr. Grubbs' Tables restrict themselves to probability levels almost universally adopted for acceptable quality level and lot tolerance per cent defective.

Mr. Clark's paper yields values of lot fraction defective, too, as its product but for probabilities which are complementary to those of the binomial tables; his probability levels are .005, .010, .025 and .050,

corresponding to binomial table probabilities of .995, .990, .975 and .950. The variable X of Mr. Clark's paper decreased by 1 corresponds to the c of the binomial tables: that is, X-1=c. As an example, entering Mr. Clark's table for n=10, and probability=.050, for X=4 (c=3) we read 1500, or .15 since Mr. Clark's tabular values are multiplied by a factor of 10,000. This is the same example, of course, as the one used in connection with Dr. Grubbs' material.

While physical examples of Gen. Simon's charts, Dr. Grubbs' tables and Mr. Clark's tables have not been displayed, it is hoped the illustrations above of their relations to the binomial tables will arouse interest in them and provide an introduction to the incomplete beta function, which has so many ramifications.

APPLICATIONS OF STATISTICAL METHODS TO THE CONTROL OF REFINERY OPERATIONS

W. Harold White, A.J. Stephenson and L.C. Greenop
Imperial Oil Limited

INTRODUCTION

It has been said that those members of a certain scientific profession who had no experience with statistical methods could be classified into three groups:

> "those who know that all statistical concepts are too difficult for their comprehension; those who know that statisticians deliberately make simple questions appear difficult in order that they may practise the mysteries of their craft; and those who know that statistical analysis of their data is unnecessary".

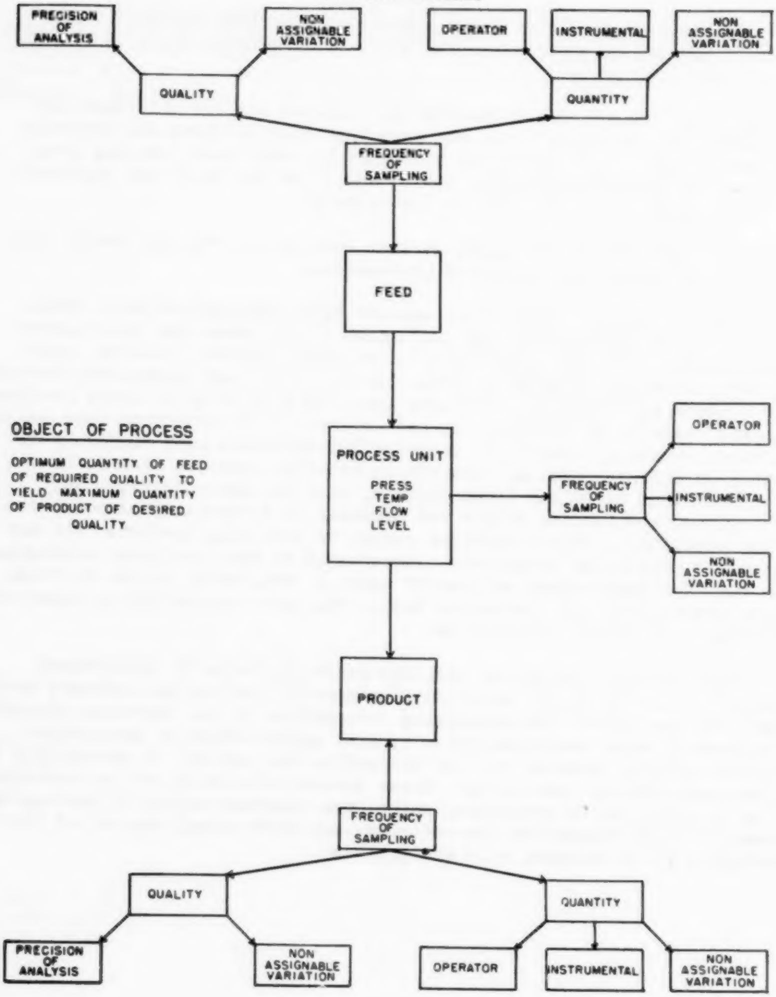
Personal observation has suggested that such groups may also exist within the engineering and chemistry professions.

Until recently, statistical methods have been applied most extensively in the agricultural and biological fields where the experimenter must usually work with materials of a variable nature. However, such variations are not limited to these two fields but are widespread throughout science and industry. The petroleum industry is by no means peculiar in this respect, as evidenced by past research and production data where, in many instances, unknown or uncontrolled variables were obviously exerting an effect. Such an observation is to be expected when it is appreciated that petroleum technologists, like the agriculturists and biologists, are dealing with a raw material of essentially unknown, complex composition, whose behaviour cannot be precisely predicted for any given set of imposed conditions. Acceptance of this statement emphasizes the need for statistical methods if correct conclusions are to be drawn from experimental and production data. The mere acquisition of numerical results is of little or no value.

Fortunately, during the past few years the value of statistical methods as a tool in the solution of research, testing and refinery problems has been given ever-increasing recognition by the petroleum industry. The present paper describes three typical applications of statistical methods made by Imperial Oil for evaluation and control of variability in petroleum refinery operations. These concern evaluation of the variability in operation of processing units; the reproducibility of testing between refinery inspection laboratories; and statistical control of the packaging of lubricating oils and greases.

FIGURE I

SOURCES OF VARIATION
IN A
REFINERY PROCESS



OBJECT OF PROCESS

OPTIMUM QUANTITY OF FEED OF REQUIRED QUALITY TO YIELD MAXIMUM QUANTITY OF PRODUCT OF DESIRED QUALITY

EVALUATION OF PROCESS VARIABILITY

The first application to be discussed concerns a general procedure for statistically evaluating the magnitude of operating variations in petroleum refining units. In general, minimizing such variability results in economic savings from improved product yields and quality.

As you may be aware, the conventional methods for evaluating the performance of a unit are to conduct special test runs under closely controlled conditions or to select and average the routine operating data. Both procedures have disadvantages in that test runs are costly and do not necessarily represent normal operation, while averaging of plant data gives no indication of their variability. The method to be described attempts to overcome these objections by using the regular operating records and then evaluating the normal fluctuations in the critical process variables, such as temperature and pressure, by statistical methods.

Basically, any refinery process may be described as a means whereby the optimum quantity of feed of the required quality is treated under given conditions of pressure, temperature, flow or level to yield the maximum quantity of products of the desired quality. Invariably, the individual components of the process are not fixed, but fluctuate due to one or more factors. The sources of variation in a refinery process, as visualized here, are shown in Figure I. From this it may be seen that variability observed in the quality of either the feed or product may be attributed to the reproducibility of the inspection test procedures and to certain non-assignable causes, which, for the present, are unknown. As an example, poor sampling technique might well be a cause of non-assignable variation. Similarly, the sources of fluctuation in the quantity of feed, of product, or in the critical process variables, such as temperature and pressure, are the indicating and control instruments, operating personnel and non-assignable causes.

It will be noted that a factor intimately associated with each of these sources of variation is the frequency of sampling. This refers not only to the number of samples of feed and product that are taken for laboratory evaluation, but also the frequency at which readings of temperature, pressure, etc. are made. Monetary and time considerations dictate that sampling should be restricted to a minimum, but, at the same time, a sufficient number should be taken to give a realistic evaluation of the variability.

The debutanizing of gasoline, a relatively simple process, will serve to illustrate the general procedure, which has been followed in the statistical evaluation of the performance of petroleum refining units. In this particular unit, cracked gasoline is fractionated in a tower to yield an overhead product containing a minimum of pentanes, which is passed to a nearby synthetic rubber plant; and a bottoms' product with a minimum of butanes, which is sent to the gasoline finishing and blending plant. Detailed examination of the operation of the unit showed that three temperatures, one pressure, three flows, and one reflux ratio might be critical with respect to product yield and quality.

As the first step in the study, it was necessary to determine the minimum frequency of taking readings for temperatures, etc. It was found that this could be accomplished by comparing the variations in readings taken at short intervals with those taken over longer periods and assess-

ing the statistical significance of any differences observed by means of an analysis of variance. For the debutanizing unit, a comparison of 7.5 and 60-minute readings showed that the latter frequency gave adequate evaluation of the fluctuations in pressures, temperatures and reflux ratios.

Determination of the appropriate sampling frequency for flow rates presents an added problem in this particular unit since they were not maintained constant, but varied rather widely depending on the number of cracking units on stream. In view of this, direct assessment of fluctuations in the volumes of flow would be of little practical value. However, it has been found that expression of each of the flows of products as a ratio of the feed rate to the fractionating tower gives a satisfactory index of their variations. It should be noted that the use of such a ratio is based on a reasonable assumption of uniformity in the composition of the feed and requires knowledge of the residence time of the various products in the unit.

It has been found that residence time can be estimated satisfactorily without the necessity of resorting to tracer techniques or other physical means if the level of bottoms' product in the tower is maintained reasonably constant. Let us consider, for example, the calculation of the average residence time of the debutanized gasoline in the unit. If values of the flow of feed to the debutanizer are plotted against the corresponding yields of debutanized gasoline, not only can the most probable relation between the two variables be estimated statistically, but also a measure of the scatter of the individual points about this line be obtained. Since the debutanized gasoline has a finite residence time in the unit, it would be expected that any changes in the flow rate of feed to the unit would be reflected at a later time in withdrawal of the bottom's product from the unit and that this interval would correspond to the residence time. In other words, if a series of regression equations are calculated between the feed rate to the debutanizer and the withdrawal of bottoms' product at various time intervals later, the degree of interrelation between the two variables should improve as the actual residence time is approached and be optimum at that point. Hence, the standard error of estimate, a measure of the scatter of the points about the regression line, would be a minimum at the residence time. Application of this technique to the debutanizer has indicated the residence time of the gasoline in the unit to be of the order of 15 minutes.

When the appropriate sampling frequencies have been determined, fluctuations in the critical operating variables can be quantitatively evaluated by simple calculations of their standard deviations. Application of the general procedure described to the operation of the debutanizing unit over a three-month period yielded results comparable to those presented in Table I. In this particular instance, the standard deviations were determined on a daily as well as total period basis to give measures of both the range and average of the operating variability. It should also be noted that, in arriving at these results, individual values exceeding three times the standard deviation were rejected and the calculations repeated. The effectiveness of statistical control is evident from a comparison with the results obtained for operation of the plant under more or less normal conditions.

TABLE I

STANDARD DEVIATIONS FOR IMPORTANT PROCESS VARIABLES IN THE OPERATION OF A GASOLINE DEBUTANIZING UNIT

PROCESS VARIABLE	SAMPLING FREQUENCY	STANDARD DEVIATION					
		NORMAL OPERATION			CONTROLLED OPERATION		
		DAILY		THREE-MONTH PERIOD	DAILY		THREE-MONTH PERIOD
		LOW	HIGH		LOW	HIGH	
FEED TEMP	1 HR	1.1	44	7.3	1.3	6.9	2.9
TOWER TOP TEMP	1 HR	1.1	37	6.5	0.83	7.5	3.3
TOWER BOTTOMS TEMP	1 HR	1.2	96	15	0.95	7.8	3.4
REFLEX RATIO	1 HR	0.04	19	3.5	0.02	4.9	0.98
GASOLINE YIELD *	1 HR	0.02	0.49	0.07	0.02	0.17	0.06
GAS YIELD *	1 HR	0.04	1.3	0.33	0.04	0.28	0.13

* EXPRESSED AS A RATIO OF THE FEED

Once having determined the magnitude of the process variations, their significance in terms of product yield and quality can be assessed. If found to be important, the overall variations, for example, in the debutanizer bottoms temperature, can be resolved into the three components mentioned previously; namely, operator, instrumental and non-assignable. In addition, by setting up control charts, marked deviations in the variables from normal can be readily detected and corrective steps taken.

REPRODUCIBILITY OF TESTING BETWEEN REFINERY INSPECTIONS LABORATORIES

Another important sphere in the control of refinery operations where statistical methods have been found of value is the inspection laboratory. As in most technical industries, the laboratory plays a vital role not only in controlling refinery processes but in assuring nationwide uniformity in product quality. The success of the laboratory in achieving these aims is due in large measure to the type and quality of the analytical procedures available. One of the most important criteria of a suitable method is that, within reasonable limits of time and material costs, it should be capable of yielding reproducible results, not only by different analysts in any one laboratory, but especially by analysts in different laboratories.

Experience has indicated that a periodic co-operative program of testing between laboratories is one of the most satisfactory ways to evaluate an analytical method or the efficacy with which it is handled in individual laboratories. While this approach is not novel, it would appear that in the past the results of some programs may have been inconclusive, inefficiently utilized or actually misinterpreted because of

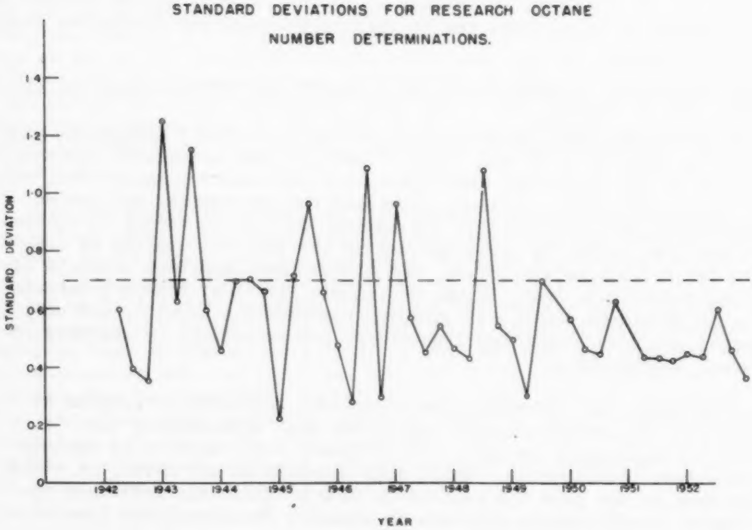
TABLE II

HISTORICAL STANDARD DEVIATIONS FOR A.S.T.M. TESTS ON GASOLINE

TEST	STANDARD DEVIATION
A.P.I. GRAVITY	0·236
SULPHUR CONTENT	0·00846
TETRAETHYL LEAD	= 0·0197 TEL + 0·0311
REID VAPOUR PRESSURE	0·633
MOTOR OCTANE NUMBER	0·594
RESEARCH OCTANE NUMBER	0·650
ACID HEAT	= 0·0649 (ACID HEAT) + 2·70
OXIDATION STABILITY	= 0·223 (BREAKDOWN) − 3·43

FIGURE II

STANDARD DEVIATIONS FOR RESEARCH OCTANE NUMBER DETERMINATIONS.



the methods used for their analyses. Such difficulties can be largely overcome by the application of some of the simple, conventional methods of statistical analysis.

The first example concerns estimation of the precision of testing between laboratories. When the results for a sample were received from all the participants in the program, they were tabulated and precision of testing evaluated by calculation of the standard deviation. However, sometimes a laboratory would report a result which appeared to be out of line with the others and should possibly have been discarded. As a working criterion, such values were rejected if they differed from the sample average by more than three times a "historical" standard deviation.

The historical standard deviation for a given analytical method has been determined, using an analysis of variance and the above criterion of rejection, on the results obtained over a ten-year period of a quarterly exchange of samples amongst six refinery laboratories. Typical values of the historical standard deviation for a number of common inspections on gasoline are given in Table II. In some instances, for example, the determination of tetraethyl lead, it was found that the reproducibility of the test increased or decreased in a regular manner with the magnitude of the measured variable. When this occurred, the correlation coefficient was calculated and, if significant, the regression equation determined.

Any consistent improvement or deterioration in the reproducibility of testing between laboratories was readily detected by plotting the standard deviations for the periodic samples against time. This is illustrated in Figure II, which shows the results obtained by several laboratories participating in a co-operative program on the determination of Research octane numbers. It will be noted that, in this instance, the reproducibility has markedly improved since 1949, which coincides approximately with the introduction of statistical control of the testing program.

When the precision of a method appears by visual inspection to have generally improved, consideration should be given to revising the historical standard deviation. This was accomplished in a quantitative manner from an "F" test on the historical variance and that of recent results. If statistically significant, the historical standard deviation required revision.

Another application of statistical methods deals with evaluation of the relative performance of various laboratories over a period of time. To make the comparison, the difference between a particular laboratory's result and the mean was noted for each sample. These differences were in turn averaged, and their standard deviation calculated. The laboratory with the average difference closest to zero has obtained the most accurate results; that with the smallest standard deviation has shown the greatest precision. The results given in Table III on the determination of Research octane number by several laboratories over a five-year period illustrates this application. It will be noted that laboratory C was the most precise and F the most accurate.

TABLE III

COMPARATIVE PRECISION IN RESEARCH OCTANE NUMBER DETERMINATION BY REFINERY LABORATORIES FOR A FIVE-YEAR PERIOD

LABORATORY	MEAN DIFFERENCE IN OCTANE NUMBER	STANDARD DEVIATION OF MEAN DIFFERENCE
A	-0.41	0.586
B	-0.33	0.619
C	-0.16	0.482
D	+0.39	0.629
E	+0.58	0.588
F	+0.06	0.532

STATISTICAL CONTROL OF THE PACKAGING OF LUBRICATING OILS AND GREASES

The final example of the application of statistical methods to refinery operations concerns control of the packaging of lubricating oils and greases. Several years ago a survey of these operations indicated not only appreciable variability but a considerable amount of product loss due to excessive overfilling of the containers. As a result, a program of statistical quantity control was introduced and has paid excellent dividends. In illustration, statistical quantity control of the filling of one-quart cans with lubricating oil reduced the average overfill about 40 percent and materially improved uniformity. In view of the large number of containers processed, this has resulted in yearly savings of several thousands of gallons of finished lubricating oils to the Company, and at the same time given added assurance that the customer received the stated amount of product.

A few of the pertinent details concerning the program may be of interest. Since it was considered impractical to empty the containers and directly determine the volume or weight of the contents, it was necessary to prepare average and range charts for both empty and filled containers. Samples, consisting of five filled and empty packages, were taken shortly after any adjustment of a machine in case of an automatic operation and after any change in operator for manual filling. The desirable frequency of subsequent sampling has been found to vary with the particular filling operation.

The control limits for the packaging of the conventional grades of lubricating oils were found to be independent of their viscosities. Hence, although the actual control of these operations has been by weight, it was possible to express the control charts in terms of volume of con-

tents and thereby avoid the necessity of preparing separate charts for products of different specific gravities. In practice, when the average weight of the empty containers in a given sample was in control, the volume was determined by subtracting the overall average weight of the empty containers from the individual members of each sample, and dividing these differences by the appropriate specific gravity factor for the product. However, if the average weight of the empty containers was out of control, this value, rather than the overall average, was used in arriving at the volume of the contents. For products filled to weight, such as greases, control charts for the filled packages were based on the total weight of the containers and their contents.

Experience with the program has indicated a few points to merit special consideration. The personnel responsible for control of the program should preferably be divorced from direct participation in the packaging operations. In our operations they are associated with the refinery inspection laboratory. It has also been found that, with any volumetric filling machine, the temperature of the product should be maintained approximately constant. Finally, it is important that the weight of the empty containers be reasonably uniform.

CONCLUSIONS

It is hoped that the foregoing has illustrated, at least in part, the advantages to be gained by using statistical methods in the control of refinery operations. While the particular examples described represent only three of numerous potentially attractive applications, their general suitability has been demonstrated by practical experience.

REPLICATION DEGENERACY

D. S. Villars
U. S. Naval Ordnance Test Station (Inyokern)

Abstract

It is pointed out that the different types of treatments applied to the ultimate sample subjected to test may be replicated to differing extents; some treatments are fully replicated, corresponding to the total number of samples tested, but other treatments may be only partially replicated. The error degrees of freedom on which the latter should be tested must be correspondingly reduced. It is important to recognize such situations in order to take account of and even capitalize upon them in the statistical analysis. Methods for identifying the correct form for the variance analysis and for deducing the appropriate significance tests are cited. To illustrate the principles, examples are discussed of both the split plot type of replication degeneracy and of a new multiple type not previously described.

Introduction

On accepting your gracious invitation to participate in this session on experimental design, it was felt that a most worth while topic for discussion would be "Replication Degeneracy". Replication degeneracy is a term introduced by the speaker (1), pp. 9, 82, 129 about fifteen years ago to designate replication to a lesser degree than is apparently indicated by the number of samples under test. It concerns a situation the industrial research man frequently encounters and constitutes the most common pitfall in the way of achieving a rigorous statistical analysis. Although the worker usually contrives to obtain a seemingly adequate number of samples for measurement, it often turns out that he cannot legitimately count all of them, but only a fraction, as true replications. In other words, the apparent degree of replication has become degenerate. Whenever conclusions from a statistical analysis violate common sense, it usually turns out that there has been an erroneous analysis of replication degeneracy in the design.

One type of replication degeneracy is well known under the term "split-plot". Another type of design which often, although not necessarily always, involves replication degeneracy has recently come to be described by the term "nesting". Analytical chemists have learned to beware of conclusions based on "duplicity" (2) of analyses (duplication of analyses without replicating treatments). This is a form of replication degeneracy. In their statistical criticism of the first Kinsey report, Cochran, Mosteller and Tukey (3) have cited the invalidity of certain standard deviations because "clustering" was not taken into account. "Clustering" introduces replication degeneracy.

Since an important type of replication degeneracy is already well known under the term, "split-plot", one might ask, "Why then introduce a new term?" There are two justifications for this. In the first place, the new term describes the basic mechanism of the phonemonen involved, thereby alerting the worker to be on guard with appropriate precautions. Of greater importance, however, is perhaps the fact that a "split-plot" type of design is a sub-case of a more general principle which the new term is desired to connote. A different but important class of designs which involve replication degeneracy but which cannot

be said to involve plot splitting will be described in the present paper.

One might think that the easiest way out of the embarrassment of having replication degeneracy in a design would be to avoid it. This is not the case for three reasons. In the first place, it sometimes happens that the nature of the problem is such that replication degeneracy is unavoidable. In the second place, and more important, cost considerations may be such that it is much more economical to deliberately design to include replication degeneracy. The economy possible with this type of design can be an important consideration during a war, not only with respect to the purely monetary aspects of cost of material wasted, but also from the point of view of conserving national resources. In the third place, it sometimes happens that our main interest is in the "sub-treatments". In such an event one can achieve a much more sensitive experiment by using replication degeneracy.

Split-Plot Type of Replication Degeneracy

In order to put across better what has been alluded to in general terms, let us elucidate by means of some concrete examples. As example one, consider the measurements of a quantity called "set-viscosity" in an experiment on rubber tread stocks. See table 1 which is taken from reference (1), p. 134. Six different formulas were mixed in duplicate,

TABLE I

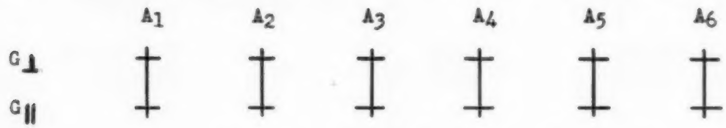
Treatment Design - Simple

	Set Viscosity		
	Mix		Total
Tread Formula	1	2	
Orthex	1550	1608	3158
Tonox-orthex	1571	1514	3085
Steam Processed	1441	1460	2901
Reclaim	1451	1385	2836
ONV	1651	1561	3212
40 Black-orthex	1576	1827	3403

but in random order. That is, a serial number was assigned to each of the twelve mixes, twelve numbers were drawn at random, the mix corresponding to the first number drawn was made up first, that corresponding to the second number drawn made up second, and so on. In making up experimental mixes of rubber, a pound lot was deemed to be a minimum size. Any amount smaller would get an abnormal degree of breakdown on a mill and the lower molecular weight produced would lead to non-representative results. In this particular experiment, it was desired to determine whether there might be a grain effect correlated with the direction of milling. Since not more than 10 grams of the cured slab of rubber was actually required for the final test, it would have been very wasteful to compound 24 pounds of rubber, if 12 could be

made to do. Accordingly, two samples were taken from the slab cured from each mix, one to be cut along the grain and the other to be cut perpendicular to it. Where replication degeneracy (split-plot type) comes into the design is at this point. Although the number of samples for test was doubled, the number of replications of the formula treatments was not. For, suppose the mill room boy had made a mistake in weighing the ingredients which went into mix 2 of formula 3. That same mistake would carry over into both the sample cut perpendicular to and the one cut parallel to the grain. So, regardless of the number of samples into which the rubber cured from any particular mix was cut, the formulas still were only replicated two-fold. The apparent four-fold replication of formulas had degenerated into two-fold. On the other hand, the grain tests (treatments) were replicated the full amount of twelve-fold, and there was no degeneracy.

The scheme of replication which was employed in this experiment may be diagrammed as in figure 1 (1 , p. 304). Each separate line represents



Figrre 1. Simple Replication Degeneracy

a single replication of the corresponding treatment. Thus there are two only for each A type treatment (formulas in the present example) but twelve for each G (grain) direction.

The individual results for the different grain directions are given in Table 2, with the analysis of variance at the bottom. A dotted line has been drawn in to warn of replication degeneracy. With this type of design there are two error estimates. The one just above the dotted line concerns the reproducibility of setting up the main treatments (formulas in this case), while that below the dotted line concerns the sub-treatments (grain direction) and interactions. In general, the former error estimate contains components of variance of the latter estimate, so is usually significantly larger. The fact that it is larger and that the number of degrees of freedom associated with it is considerably smaller leads to the requirement of an appreciably greater critical F ratio to demonstrate significance. Hence, introduction of replication degeneracy leads to a less sensitive experiment with respect to the main effects, but this is compensated for by the fact that the test is usually more sensitive with respect to the sub-treatments and their interactions. Because of these facts, it is therefore always well to map out the analysis of variance before preparing any of the treatments. If it should be indicated that not enough error degrees of freedom will be available, it is then not too late to redesign the experiment so as to provide an adequate number.

For any particular design there is only one correct analysis of variance. The lower analysis in Table 3 shows a possible alternative form which might be erroneously chosen by an inexperienced designer. The alternative analysis would have been correct had there been made twenty four mixes at random, one each to be used for each different

TABLE II

Treatment Design - Factorial

Classification by Formulas, Direction of Force Relative to Grain, and by Times of Set

Tread Formula	Set Viscosity Mix 1		Mix 2		Total	
Direction to Grain	‖	⊥	‖	⊥		
Orthex	798	752	786	822	3158	Formula Classes
Tonox-orthex	775	796	764	750	3085	
Steam Processed	734	707	734	726	2901	
Reclaim	732	719	707	678	2836	
ONV	838	813	753	808	3212	
40 Black-orthex	802	774	904	923	3403	
					18595	

‖ Total	9327	Direction Classes
⊥ Total	9268	

Variance Analysis

Classification	S.Sq. Dev.	D.F.	Variance Estimate	Sig.
Between Formulas, F	54309	5	10862	?
Within Formulas, W	20608	6	3435	
Between Directions, D	145	1	145	-
Interactions DF	885	5	177	-
Error DW	4293	6	715.5	

TABLE III

Replication Degeneracy - Correct Variance Analysis

	D.F.
Between Blocks, B	3
Between Main Treatments, M	3
Error, MB	9
Between Sub-treatments, S	2
Interaction, SM	6
Error, S (B + MB)	24
Total	47

Table III (Cont'd)

Incorrect Variance Analysis

Between Blocks, B	3
Between Main Tests, M	3
Between Sub-tests , S	2
Interactions, SM	6
Error	33
Total	47

(Note: The main treatments, M, may only be tested on a particular set of 9 error d.f.--it is very wrong to test them on the 33 in the last analysis. Significant results from the latter would be entirely misleading.)

Table IV

Multiple Splitting

Treatments		Degrees Freedom	Mean Square	Variance Ratio	Significance	Variance Component
Between Fabric Treatments	F	2	6.5749	21.55e	+++	.0159(+++)
Within " "	W	12	0.3043e			
Between Rubber Treatments	R	1	0.3147	18.95dd	+++	(but ν_R -)
Interaction	RF	2	0.3115	18.75dd	+++	
Rubber Treatment Error	RW	12	0.0166d			$\nu_{RW} < 1$
Between Cures	C	1	0.7492	14.88 c		
Interaction	CF	2	0.1003	2.050c	-	
Interaction	CR	1	0.1237	2.460c	-	
Curing Error	CRF	2				
	CW	12	0.0503c			ν_{CW} = .00235(-)
	CRW	12				
Between Test Temp.	T	1	0.9614	23.50 b	+++	(but ν_T -)
Interaction	TF	2	1.4392	35.20 b	+++	
	TC	1	0.0705	1.723b	-	
	TCF	2	0.0227	0.555b	-	
Test Temp. Error	TR	1				
	TRF	2				
	TCR	1				
	$TCRF$	2	0.0409b			ν_{TW} = .0169(+++)
	TW	12				
	TRW	12				
	TCW	12				
	$TCRW$	12				
Testing Error	E	120	0.0072a			ν_E = .0072

grain-formula combination (as diagrammed in Figure 2). If one should carry out the analysis of variance

	A_1	A_2	A_3	A_4	A_5	A_6
$G_{\perp}$	++	++	++	++	++	++
$G_{\parallel}$	++	++	++	++	++	++

Figure 2. Complete Randomization

for the design of Figure 1 using the lower scheme of Table 3 (appropriate only for Figure 2), he would get the same values of mean squares for B, M, S and SM. The only difference would be that only one error estimate would be obtained and it would be a weighted average of MB and S(B + MB). So if one should test M on this average error, which would be smaller than the correct error, MB, with 33 degrees of freedom instead of the correct number, 9, he could get an indication of high statistical significance when there was none whatever. Conversely, a test of sub-treatments, S, and interactions, SM, on the average error would tend to be over-conservative. In either case, the technique would be erroneous.

One is not limited to a single sub-division of sample in utilizing the principle of replication degeneracy. As example 2 let us take a more complicated factorial design employing multiple splitting (see Table 4 taken from ref. 1, p. 136 ff.) This design was set up by Mr. Joseph Rouark of the U. S. Rubber Co. Three different tire fabrics were made up each replicated five-fold. The 15 lots were then subdivided into two halves and two different types of rubber treatment were applied to each half. Each piece of the resulting fabric was then further sub-divided into halves, one half being subjected to one curing procedure, the other half to a second type of cure. The 60 samples resulting were then further sub-divided into two each, one to be maintained at room temperature and the other at an elevated temperature in an oven. Tests were finally run in duplicate. At each point of subdivision in the table, a dotted line is drawn to warn against replication degeneracy. Each effect is tested only on the error estimate within its own section of the table as designated by the dotted lines.

Application of replication degeneracy is not restricted to factorial designs. Sub-division of test material may be introduced at any stage of any design whatever.

<u>Multiple Replication Degeneracy</u>

At the beginning of this paper reference was made to another type of replication degeneracy which cannot be classified as a split-plot design. This involves a multiple degeneracy (1), p. 308.. As example 3, let us suppose that while evaluating two different polymers one might be simultaneously interested in whether the product prepared early in the week, just after the week end, differed in quality from that manufactured in the middle of the week. The designer, knowing about replication degeneracy, does not fall into the pitfall of making up only one independent batch of each polymer, but makes up, say two

each. He then sets up the experiment so that the four batches, two of polymer A_1 and two of polymer A_2, are tested on a certain Monday. From the same four batches, further samples are prepared and tested again on the following Wednesday. The replication scheme may be diagrammed as in Figure 3.

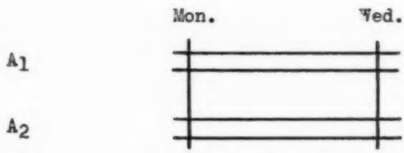


Figure 3. Double Replication Degeneracy

It seems intuitively clear that inferences deducible from a design set up according to Figure 3 could be quite different from the inferences from a design in which a different Monday and a different Wednesday were used for each of the four batches (diagram in Figure 4) or from a design where all tests were run on a single Monday and a single Wednesday but four separate batches were made up of each polymer (replication diagram in Figure 5). For, only in the next to last alternative (Figure 4) could one expect to infer a drop in quality on Mondays to represent

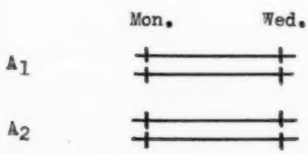


Figure 4. Alternative Simple Replication Degeneracy

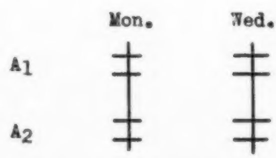


Figure 5. Alternative Simple Replication Degeneracy

support of a hypothesis of a "week-end hang-over" effect. In the other designs, although a low value for Monday might be statistically significant, it might have arisen from special conditions prevailing on that date, such as temperature or humidity, and not be characteristic of Mondays in general.

Determination of Correct Form of Analysis of Variance

The preceding complications of degeneracy may be multiplied, once for each factor in a factorial experiment. Diagrams are not given for higher orders of degeneracy because they would entail more than two dimensions on paper. In view of this manifold of complications, how can a humble worker ever know what is the correct form of analysis of variance for the particular design he has chosen? The writer has shown (1), p. 304 that the correct form of analysis may be very simply deduced by writing the so-called mathematical model of all the different possible sources of variation which it is desired to take into consideration in the experimental measurement. This sounds highly complicated but actually is fairly simple. Models for the designs of the preceding figures are summarized in the following equations:

Figure	Equation
1	$x_{ijk} = \xi_o + \alpha_i + \delta\alpha_{ij} + \beta_k + \delta\beta_{ijk} + \iota_{ik} + \delta\iota_{ijk}$
2	$x_{ijk} = \xi_o + \alpha_i + \delta\alpha_{ijk} + \beta_k + \delta\beta_{ijk} + \iota_{ik} + \delta\iota_{ijk}$
3	$x_{ijk} = \xi_o + \alpha_i + \delta\alpha_{ij} + \beta_k + \delta\beta_k + \iota_{ik} + \delta\iota_{ijk}$
4	$x_{ijk} = \xi_o + \alpha_i + \delta\alpha_{ij} + \beta_k + \delta\beta_{ijk} + \iota_{ik} + \delta\iota_{ijk}$ (same as Fig. 1)
5	$x_{ijk} = \xi_o + \alpha_i + \delta\alpha_{ijk} + \beta_k + \delta\beta_k + \iota_{ik} + \delta\iota_{ijk}$

In these equations single Greek letters are used to represent the true effects of treatments to be considered, while those with a δ in front represent the error made in setting up that treatment in the experiment. ξ_o is the true value one would have obtained for the measurement in the absence of any treatment effect or error. The subscripts refer to which treatment, i to the A's (formula or polymer in our examples), j to the batch, and k to the B's (grain directions or days in these examples). Extreme care has to be taken to be sure that the subscripts correctly represent the effect being represented. The reader is invited to study the why of the different arrangements in the above model equations. Once the subscripts have been assigned, all terms with the same set of subscripts are grouped together, inasmuch as all effects determined by the same subscripts will be indistinguishable from one another. The correct variance analysis will then contain one line for each group. Any line representing a group containing an error component (a δ) will be worthless for inferences concerning a main effect or interaction, since a significant result could be attributable to such an error component and the result would accordingly be ambiguous. Thus, in the designs for Figures 3 and 5 there is no way to demonstrate a B effect, since it would be confounded with the non-reproducibility in setting up the B treatment. It may be seen that there will be respectively 5, 4, 5, 5 and 4 lines in the designs corresponding to the above figures.

Determination of Appropriate Significance Test by Components of Variance

In order to determine which F ratio to use for the correct significance test, the most clear-cut way is to resort to a table of components of variance (1, Chapt. 7). For example, Table 5 shows that formation of

the ratio of mean squares, F: E, would give

$$F/E = \{\nu_E + c\,\nu_{WR} + rc\,\nu_W + rcw\,[f/(f-1)]\,\nu_F\}/\nu_E$$

$$= 1 + c\,\nu_{WR}/\nu_E + rc\,\nu_W/\nu_E + rcw\,[f/(f-1)]\nu_F/\nu_E$$

A value of this ratio found to be significantly greater than unity could arise from significant effects in ν_W and in ν_{WR} as well as in ν_F, which is the effect under test. Such a ratio would be inappropriate as a significance test as it would be ambiguous. Similarly, the ratio F/WR simplifies to

$$F/WR = 1 + rc\,\nu_W/\{\nu_E + c\,\nu_{WR}\}$$
$$+ rcw\,[f/(f-1)]\,\nu_F/\{\nu_E + c\,\nu_{WR}\}$$

It would likewise be ambiguous as a significance test, as values significantly greater than unity could arise from a significant effect in ν_W as well as in ν_F. Only the ratio, F/W, which simplifies to

$$F/W = 1 + rcw\,[f/(f-1)]\,\nu_F/\{\nu_E + c\,\nu_{WR} + rc\,\nu_W\}$$

will give an unambiguous test, since only that ratio gives unity plus a single term with only the effect under test in the numerator.

Before an experimentalist can be in a position to use a table of components of variance it is first necessary for him to settle in his own mind certain postulates as to the kind of population to which each of his effects (Greek letter groups having the same sets of subscripts in equations 1 to 5 above) is presumed to belong; i.e., whether they represent the complete population or just a random sample of a much larger population. Different tables apply for different postulates. If the treatments in a certain classification represent all the treatments of that kind in which one expects to be interested, then they represent the complete population. Thus, in example 1 of this paper, the two grain directions are the only possible treatments of that classification-- any intermediate directions are made up of varying fractions of these two orthogonal components. On the other hand, Days of the week in example 3 are definitely samples of a large possible population of Mondays and Wednesdays. Sometimes this criterion may be equivocal. Thus, in the first example one could equally say that he is interested in these six tread stock formulas only or that he wishes to consider these formulas as representatives of a larger population of possible improvements yet to be found. One is free to accept either or both alternatives, provided he recognizes that each postulate sets up a different problem answers to which may be inconsistent with each other.

In Table 5 is reproduced components of variance which have been compiled by the speaker (1), p. 133 for the following four extreme cases of a three factor design with double splitting (in all cases, W and its interactions are error components and hence are to be considered random samples):

Table V

Components of Variance

Three Factors with Replication Degeneracy

Source	**D.f.**	**Components of Variance**
	Case (a) ***F, R, C* - Complete Populations**	
Bet. *f* tmts., *F*	$f-1$	$v_E + c\,v_{WR} + rc\,v_W + rcw[f/(f-1)]v_F$
Within, *W*(*w*-fold)	$f(w-1)$	$v_E + c\,v_{WR} + rc\,v_W$
Bet. *r* tmts., *R*	$r-1$	$v_E + c\,v_{WR} + fcw[r/(r-1)]v_R$
Interactions, *FR*	$(r-1)(f-1)$	$v_E + c\,v_{WR} + cw[f/(f-1)][r/(r-1)]v_{FR}$
Error, *WR*	$f(w-1)(r-1)$	$v_E + c\,v_{WR}$
Bet. *c* tmts., *C*	$c-1$	$v_E + frw[c/(c-1)]v_C$
Interactions, *FC*	$(c-1)(f-1)$	$v_E + wr[f/(f-1)][c/(c-1)]v_{FC}$
Interactions, *RC*	$(c-1)(r-1)$	$v_E + fw[r/(r-1)][c/(c-1)]v_{RC}$
Interactions, *FRC*	$(c-1)(r-1)(f-1)$	$v_E + w[f/(f-1)][r/(r-1)][c/(c-1)]v_{FRC}$
Error, *E* = (*W* + *WR*)*C*		v_E
	Case (b) ***F, R* - Complete Populations** ***C* - Random Sample**	
F		$v_E + c\,v_{WR} + rc\,v_W + rw[f/(f-1)]v_{FC} + rcw[f/(f-1)]v_F$
W		$v_E + c\,v_{WR} + rc\,v_W$
R		$v_E + c\,v_{WR} + fw[r/(r-1)]v_{RC} + fcw[r/(r-1)]v_R$
FR		$v_E + c\,v_{WR} + w[f/(f-1)][r/(r-1)]v_{FRC}$ $+ cw[f/(f-1)][r/(r-1)]v_{FR}$
WR		$v_E + c\,v_{WR}$
C		$v_E + frw\,v_C$
FC		$v_E + wr[f/(f-1)]v_{FC}$
RC		$v_E + fw[r/(r-1)]v_{RC}$
FRC		$v_E + w[f/(f-1)][r/(r-1)]v_{FRC}$
E		v_E
	Case (b') ***F* - Complete Population** ***R, C* - Random Samples**	
F		$v_E + c\,v_{WR} + rc\,v_W + w[f/(f-1)]v_{FRC} + rw[f/(f-1)]v_{FC}$ $+ cw[f/(f-1)]v_{FR} + rcw[f/(f-1)]v_F$
W		$v_E + c\,v_{WR} + rc\,v_W$
R		$v_E + c\,v_{WR} + fw\,v_{RC} + fcw\,v_R$
FR		$v_E + c\,v_{WR} + w[f/(f-1)]v_{FRC} + cw[f/(f-1)]v_{FR}$
WR		$v_E + c\,v_{WR}$
C		$v_E + fw\,v_{RC} + frw\,v_C$
FC		$v_E + w[f/(f-1)]v_{FRC} + wr[f/(f-1)]v_{FC}$
RC		$v_E + fw\,v_{RC}$
FRC		$v_E + w[f/(f-1)]v_{FRC}$
E		v_E

Table V (Cont'd)

Case (c)
***F, R, C* - Random Samples**

F	$v_E + o v_{WR} + ro v_W + w v_{FRC} + rw v_{FC} + ow v_{FR} + row v_F$
W	$v_E + o v_{WR} + ro v_W$
R	$v_E + o v_{WR} + w v_{FRC} + fw v_{RC} + ow v_{FR} + fow v_R$
FR	$v_E + o v_{WR} + w v_{FRC} + ow v_{FR}$
WR	$v_E + o v_{WR}$
C	$v_E + w v_{FRC} + fw v_{RC} + wr v_{FC} + fwr v_C$
FC	$v_E + w v_{FRC} + wr v_{FC}$
RC	$v_E + w v_{FRC} + fw v_{RC}$
FRC	$v_E + w v_{FRC}$
E	v_E

Case	Postulates
a	Treatments F, R and C represent their complete populations.
b	Treatments F and R represent their complete populations; Treatment C represents a small random sample from a large population.
b'	Treatment F represents the complete population; Treatments R and C represent small random samples from large populations.
c	Treatments F, R and C all represent small random samples from large populations.

(Case a is also known as a model I experiment, case c a model II, and the other two cases as a mixed model experiment (4). Intermediate cases have been discussed by Crump (5). Which components of variance are present or absent for any type of effect may be written down by inspection by means of the following short-cut rule which the speaker has found to work for all factorial designs so far computed by him (1), p. 121.. First, include a component for the main effect or interaction under consideration; second, include components for <u>all</u> interactions with the main effect or interaction; finally, strike out those of the latter, where one of the effects other than the main effect or interaction under consideration involves a complete population.*

---

*A different short-cut rule for inducing components of variance is cited by H. Scheffe' (6). His rule, however, does not omit those components established as being absent by the method used by the speaker.

References

(1) D. S. Villars, "Statistical Design and Analysis of Experiments for Development Research" William C. Brown, Dubuque, Iowa (1951).

(2) Term attributed to J. W. Tukey.

(3) W. G. Cochran, F. Mosteller, J. W. Tukey, "Statistical Problems of the Kinsey Report" J. Am. Statis. Assn. 48, 673-716 (1953).

(4) Churchill Eisenhart, "The Assumptions Underlying the Analysis of Variance" Biometrics 3, 1-21 (1947).

(5) S. L. Crump, "The Present Status of Variance Component Analysis" ibid 7, 1-16 (1951).

(6) H. Scheffe', "On the General Theory of Experiments Designed to Evaluate Several Sets of Constants and Several Sources of Variability", Chem. Eng. Progress, in press.

SOME APPLICATIONS OF QUALITY CONTROL TECHNIQUES TO CLERICAL WORK

Robert B. Selover
The Prudential Insurance Company of America

Introduction

In this paper I will describe two applications we have made of quality control to clerical operations. These examples will, I believe, demonstrate that there are applications of quality control techniques to clerical work as well as to industrial production. Our experience convinces us that very real gains can be made in both the quality and economy of our work by application of quite common concepts and methods. In discussing these two examples of our work I would like to point up some of the psychological principles used to encourage improvement. Effective applications of principles of learning and motivation have been made in these programs. Since these concepts have not been discussed previously I would like to emphasize them in this report.

Development of Program

Dr. Bennet Murdock first made application in our Company of quality control techniques in June, 1949. This first attempt, applied to work involved in making changes in life insurance policies requested by the insured, was so successful that many further programs followed. We have operated programs in 35 different work groups in 12 different Divisions of the Company. The applications have covered a variety of clerical work from simple repetitive tasks such as filing, to complex work such as underwriting and correspondence. Almost without exception improvement has been striking at first, followed by more gradual improvement. A graph of the percent of cases handled correctly appears much like a typical learning curve which improves rapidly at first followed by a more gradual increase. Much of what we do is to apply principles and methods which will facilitate learning.

Methods of Approach

Although each application of quality control is different because of the nature of the work done, perhaps I can best explain how we make an installation by giving an example. There are several differences between our approach and the more commonly used methods in industrial quality control. First of all, we are primarily interested in improving the performance of people rather than controlling the quality of machines. It is for this reason that we refer to our work as quality improvement. Clerical work is either right or wrong and therefore we deal with attributes rather than variables. Mechanical means are generally not available to determine clerical accuracy and therefore we place great reliance on the inspector's knowledge and judgment. The inspection of work is made by one of the clerks, whom we call a Quality Reviewer, rather than by an outsider. We realize, of course, that disadvantages can result from this procedure but we feel that the advantages we gain are more important. Quality Reviewers are changed periodically so that each person in the section acts as a reviewer. Thus, each clerk has the opportunity to see the work of other members of the section and can learn from this experience. Employees frequently become more personally involved in the goal of improving the quality of work after they have had experience as a reviewer.

Chart 1 shows schematically how a program might operate in a typical section.

Work comes into the section and is distributed to the clerical group. After the work is completed it is sent to a control desk. A representative sample of the work is taken from the control desk by the Quality Reviewer. This sample of work is then carefully checked for accuracy and completeness. A tally is made of the number and types of inaccuracies found. Error cases are sent to the Supervisor who returns them for correction to the clerks who made the errors. The correct cases are returned to the control desk from which they leave the Division along with the rest of the completed work. The error information is sent to the Quality Improvement Section where it is analyzed to determine significant changes which have occurred in the accuracy of the work and to determine, insofar as possible, the causes for the inaccuracies. If systematic errors are observed, the Quality Improvement Staff attempts to discover whether or not some part of the processes, forms, source material or method operates to facilitate making the errors. Each week the results are tabulated by the Quality Improvement Section and distributed to the Management in the Division. Periodic reports containing analyses and suggestions are also prepared and sent to interested individuals. Each week a quality graph is prepared for distribution. This tells the story of improved quality at a glance, even though it may oversimplify progress. Each installation is, of course, different because of the nature of the work done and the methods used in processing the work.

Results

Typical results of a program are shown in Chart 2. This program was carried out on a simple card filing operation.[1] The cards in this file are used for reference to basic information about life insurance policies. A card is removed from the file and sent to another division when a change in the policy is requested. When the cards are returned they must, of course, be refiled. It is the accuracy of this refiling operation that is represented by the graph. Approximately 40,000 cards are refiled each week. The work is done as a part-time job by about 60 clerical employees each afternoon. The percent of cards filed correctly started at 98% and improved to about 99.4% over a period of half a year. The question arises as to the importance of this improvement to the Division's operation. Since this was only one of several programs of quality improvement carried out in the Division, an exact assessment of its contribution to the improvement is difficult. One indication is given, however, by the amount of time required to locate cards reported missing from the file. When a card cannot be located a special search is, of course, necessary. Before the Quality Improvement program was installed, this tracing operation required special personnel as well as supervisory time. Since the beginning of the program the amount of time spent in this tracing operation has reduced by about one half. This program was discontinued after seven months' operation because it appeared that further improvement should not be expected. A periodical review is made of the filing in order to assure ourselves of the continuation of the improved accuracy obtained by the program.

The second program I would like to mention is also on a filing operation. In this case we are interested in the accuracy with which billing cards are picked in order to audit premium receipts received from our branch offices. These billing cards are punch cards which contain identifying information and the amount of premium which is due. The billing

1 This program and the one following were planned and supervised by Joseph Quade and Blair Olmstead respectively, who are members of the Quality Improvement Section.

cards picked must agree exactly with the receipts received in order to prove the report.

After the cards have been picked they are sent to the Machine Room where a tabulator listing is made showing each premium due. Totals are shown for each district. This report is then returned to an Audit Clerk who compares this listing with the branch office report. If they do not agree, the Audit Clerk must determine why and arrange for the errors to be corrected. After an error is found, the correct billing cards are sent to the Machine Room and the report "rerun."

The program in this section was started on March 13, 1952. A summary of the results are shown in Chart 3. The error rate at that time was .7 of 1%. At the present time the error rate is fluctuating between .1 and .2 of 1%. It is interesting to note the shape of the graph of improvement. One might expect that changes within such a small range of accuracy would be haphazard and perhaps defy explanation. It is interesting that each year there is a very large change of personnel in this section. We employ high school graduates during June and July and assign many to this work. People who have been in this section are moved on to other jobs. The quality of performance shows a very noticeable drop in August each year, which is apparently related to this change in personnel. Over the period of two years a significant improvement is evident. Although the change in the percent of cases handled correctly was small, it had a very marked effect on the number of reruns necessary in the Machine Room. The number of reruns required to prove the accounts has been reduced from about 40% in 1952 to slightly under 20% at the present time. This, of course, has greatly reduced time in the Machine Room to perform the operations.

Discussion

There are several factors which I believe operate in producing the improvement we characteristically observe in these programs. These might be divided into two general classes. First, changes are made in methods and procedure which reduces the possibility of making an error. Second, principles of learning and motivation are introduced which encourage improvement.

In the first example presented, methods changes included improved physical condition of the file. An improved use of card notching was introduced which made it quite apparent when a card was misfiled in the wrong drawer. A more adequate presort was also introduced before the cards were refiled. In the second example, methods changes included improving the legibility of the billing card by changing the color border used, and improving the ink used in interpreting cards. The file itself was moved to a better location and the lighting and space improved. These changes, we believe, had a real effect in reducing the possibility of error in processing the work.

The principles of learning introduced in these programs are perhaps less readily identified but nevertheless important. First, just getting objective about quality, that is, defining it so that it can be counted, has an important effect. Quality then becomes something which can be observed, measured and plotted, rather than a nebulous, ill-defined but desirable characteristic. By measuring quality of performance employees are shown in a very real way that Management is genuinely interested in their producing work of high quality. If employees do not have that understanding we can hardly expect them to be greatly concerned about

their accuracy.

Learning is greatly facilitated by clear-cut goals toward which to work. Intermediate or sub-goals which can be attained within a reasonable time are more effective in producing learning than distant and difficult goals. Measuring quality makes it possible to set up objectives which are within easy reach of employees and thus encourages improvement.

The importance of knowledge of results has been demonstrated in all kinds of learning situations. Studies have shown that in the absence of knowledge of results, extended practice produces little, if any, improvement. Further, it is found that immediate knowledge is most effective in facilitating improvement. As the time interval is lengthened between performing an operation and information about its correctness, learning becomes more difficult. Studies have also shown that the information given should be as definite and complete as possible. These basic concepts are encouraged by a program which reports quickly and accurately the kinds of errors that have been made. You perhaps have heard a Supervisor say, "I told them exactly how many errors they made and they haven't improved in the least." Of course, one should not expect improvement in the absence of exact information about what was done incorrectly. Employees need to be given knowledge of their performance in exact terms. Results should be furnished quickly and automatically to be most effective in improving performance.

The results of quality are measured in terms of percent of cases handled correctly rather than a count of incorrect cases. This, we feel, emphasizes accomplishment and gives people an opportunity to develop pride in their work. No permanent individual records are kept, and by the way a program is installed and operated it is made clear that our objective is improving the quality of the work and not the evaluation of people. All of these factors we believe operate to produce the characteristic result we have observed.

Summary

On the basis of our experience in applying quality control techniques to clerical work, we have learned to expect a substantial increase in the percent of cases handled correctly. A typical curve of improvement is negatively accelerated, that is, it shows first a sharp increase followed by a period of more gradual improvement. This increase, we believe, is brought about by both methods improvements and the use of procedures which facilitate learning and increase motivation. Some of the more important principles used in this approach which facilitate learning are:

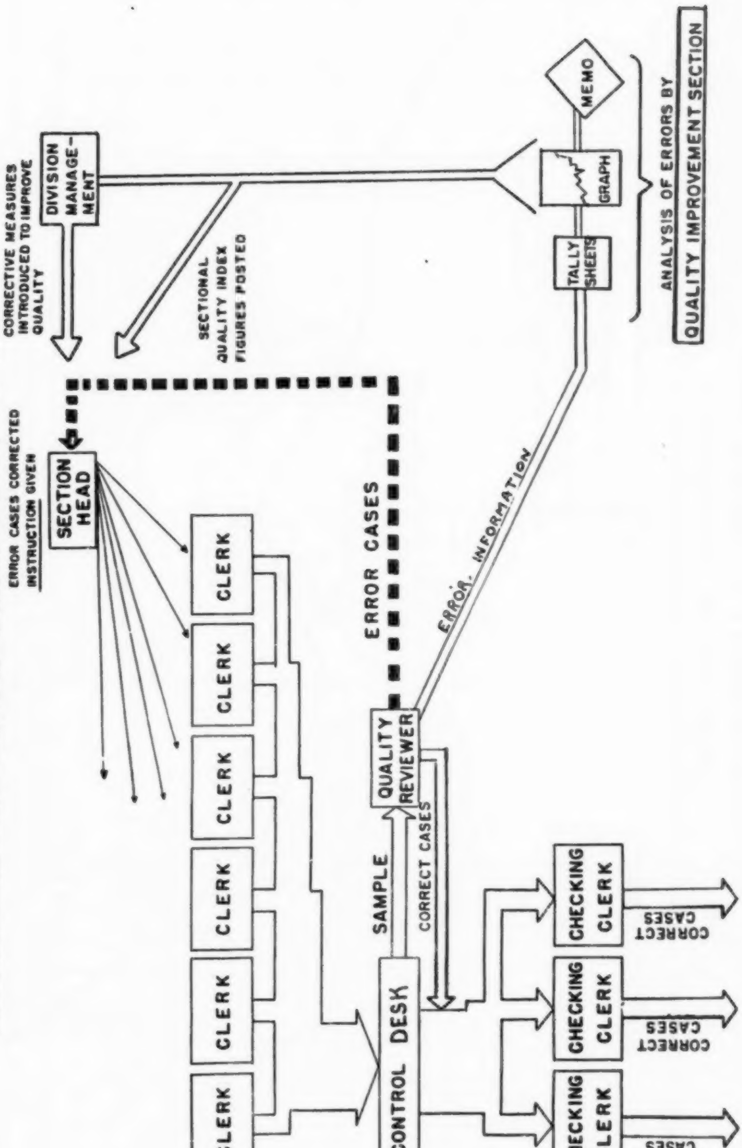
1. measuring performance objectively so that meaningful goals can be established toward which to work,

2. giving immediate and exact knowledge of results of their performance to the clerical staff, and

3. expressing results positively, thus giving people encouragement to improve and take pride in the quality of their work.

CHART I

SECTIONAL OPERATION OF A QUALITY IMPROVEMENT PROGRAM



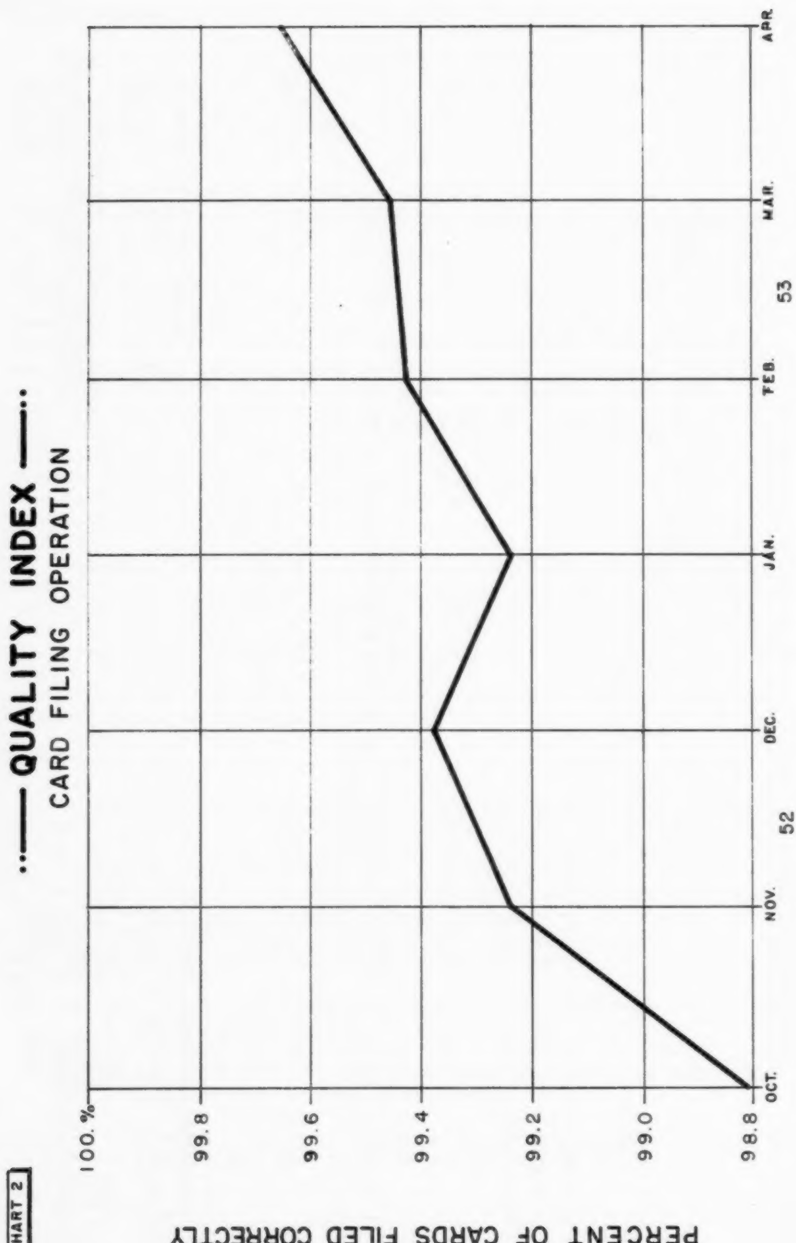


CHART 2
QUALITY INDEX
CARD FILING OPERATION
PERCENT OF CARDS FILED CORRECTLY
100. %
99.8
99.6
99.4
99.2
99.0
98.8
OCT.
NOV.
52
DEC.
JAN.
FEB.
53
MAR.
APR.

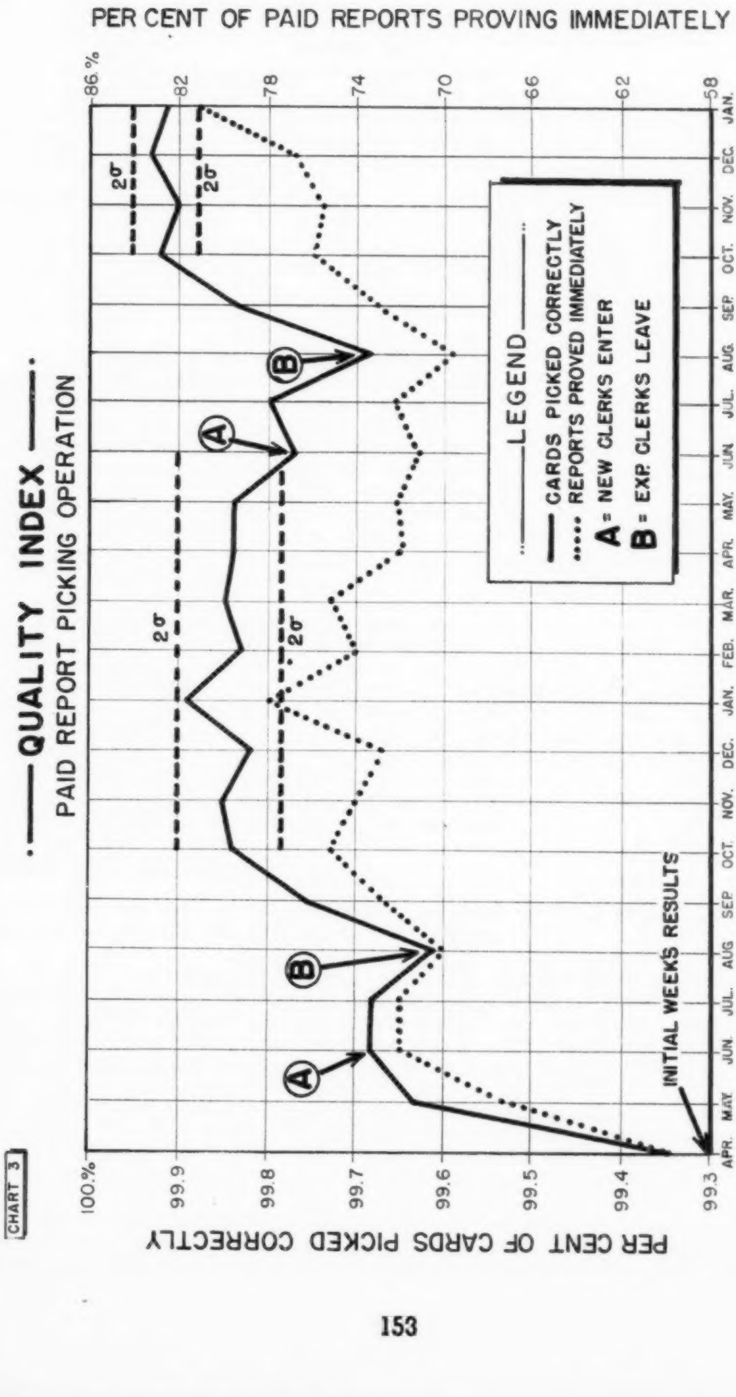


CHART 3
QUALITY INDEX
PAID REPORT PICKING OPERATION
PER CENT OF PAID REPORTS PROVING IMMEDIATELY
PER CENT OF CARDS PICKED CORRECTLY
100.%
99.9
99.8
99.7
99.6
99.5
99.4
99.3
86.%
82
78
74
70
66
62
58
APR. MAY JUN. JUL. AUG. SEP. OCT. NOV. DEC. JAN. FEB. MAR. APR. MAY. JUN. JUL. AUG. SEP. OCT. NOV. DEC. JAN.
1952
1953
54
INITIAL WEEK'S RESULTS
2σ
LEGEND
CARDS PICKED CORRECTLY
REPORTS PROVED IMMEDIATELY
A = NEW CLERKS ENTER
B = EXP. CLERKS LEAVE

QUALITY CONTROL OF COMPLEX ASSEMBLIES

Paul A. Robert
International Business Machines Corp.

The use of formal techniques for controlling the quality of finished machines is constantly increasing, and represents one of the large fields of application for the statistical approach.

CONSUMER QUALITY

In controlling the quality of a finished product, a new element is introduced. Up to the time a product is complete, it is customary to think of quality in terms of conformance to an established specification rather than the "end use" or consumer idea of quality.

The consumer's idea of quality may be quite different from that expressed by the engineer. Customers are likely to think in terms of the inconvenience caused by a defect, rather than the seriousness of the defect itself. For example, if a relay point fails to make properly, and an accounting machine produces a wrong answer, the trouble may be very minor; one that could be corrected in a minute. However, if the error causes the user to have to rerun a large quantity of cards, and delayed his work schedule, he is very unhappy and the trouble is not minor, but a very serious one. This same thinking on the part of the consumer applies whether we are talking about accounting machines, automobiles, furniture, shoes, refrigerators, or TV sets.

According to Mr. Ephraim Freedman, who is the Director of the Bureau of Standards of R. H. Macy Co. in New York City, some products have a customer complaint rate of as high as 25% of all sales. In his address before the Rochester Clinic this Spring, he painted a pretty sorry picture of the quality of the goods being manufactured in some factories. And, he went further to point out that he thought that this was due in a large measure to the fact that we do not often think in terms of the consumer. It is important that we do not restrict ourselves to compliance with specification and shop ideas of good workmanship, but be comprehensive and visualize the reaction of the consumer.

OBJECTIVES OF QUALITY CONTROL

It would be well to take a moment to review just why we are controlling and evaluating the quality of our product.

First, and foremost, of course, is to prevent the shipment of substandard items to the customer. Any effective control technique should setup a barrier against the continued shipment of "more than the desired" percentage of defective articles. And, as we shall later see, some techniques are more effective than others and operate more promptly.

Second, since it is usually uneconomical to manufacture defective material, and then correct it by 100% inspection and rework, the quality control technique points up those defects which, because of their critical nature and/or frequency of occurrence, warrant corrective action.

Third, for the proper overall management of an operation, it is important for the plant executive to have quality performance figures fully as much as figures on cost and production. Production, cost, and quality

are interrelated, and for proper administration of these functions, the plant manager requires factual measures of each.

Finally, quality reports also serve to supervise the inspection function itself. Just as inspection reports reflect the quality produced by the operating personnel, so can they be made to provide a means of evaluating the effectiveness of the inspection operation itself.

TYPES OF QUALITY CONTROL PLANS

In general, there are three types of quality control plans.

First, and probably most common, are the final inspection plans. Under this type of plan, all or a sample of the product is examined by inspectors, tested, and on the basis of their findings, the product is either accepted or rejected.

Second, a sample of the product is taken to a laboratory and tested under aggravated conditions to simulate long and arduous use. This performance test is usually made until failure occurs,and is planned to disclose design as well as manufacturing weaknesses. These tests often include simulated shipment to the customer, and tests under environments of extreme heat, coldness, pressure, moisture, etc.

Third, a survey of field performance as measured by failures, complaints, service time, consumer reaction studies, returned equipment,etc.

In all types of quality evaluation it is not possible to arrive at a firm and absolute measure of the intangible thing we choose to call quality. As we have indicated, it means different things to different people. As a result, we select what seems to be the most appropriate indicators or measures of the characteristic.

ADVANTAGES AND DISADVANTAGES

Each of these quality plans have merit, and generally all three plans are employed. Let us consider for a moment some of the advantages and disadvantages of these plans.

The final inspection plan obviously gives us the earliest indication that something is wrong. As inspectors have been trained to check adjustments, good workmanship, etc., they are sensitive to conditions which may be latent sources of trouble. Thus, at this stage we have an indirect means of controlling the long term quality, as well as the present performance.

Also, at final inspection it is possible to cover a large number of characteristics and units with a small number of people, and thus develop in some respects a more reliable measure of quality.

On the other hand, inspectors do not necessarily look at a product as a customer would. They see defects a customer would never see,but may discount others as trivial which a customer would regard as important. By careful planning and training, a formal inspection procedure can partially overcome this discrepancy. Also at final inspection, many defects remain hidden and inaccessible, and show up only after the product has been in service awhile.

In performance testing, the quality evaluation is more objective and scientific. The item under test is under observation longer, and latent

defects have an opportunity to show up. The evaluation is usually made by highly skilled and trained engineers with a point of view closer to that of the consumer. Each failure can be laboratory analyzed to determine the exact cause.

However, this type of test is usually expensive, and the very small sample may not be at all representative of production in general. The test is probably so time consuming that the results are not available until months after the test is started.

It might be said, "that the proof of the pudding is in the eating", and therefore, the field test of a product is a measure of its true quality. It is most likely to reflect the consumer's point of view. Properly designed, the field test can reflect a cross section of uses to which the product is applied.

Unfortunately, it too has its disadvantages. Products are sometimes misused by the consumer, and improperly analyzed and repaired by the serviceman. There are so many customers and servicemen that it is difficult to train such a large group to diagnose the trouble.

Where many defects are apparent on installation, many are not, and the time that elapses may be quite long because the conditions of use may not tax the equipment in a way to bring out the defect.

Ordinarily, this type of test must be considered a long range test, and the problems of administering it can be substantial. If the field organization is closely integrated with the parent organization, this can be an effective measure of "true" quality.

FINAL INSPECTION PLANS

In developing a final inspection plan, the first step is to develop a "laundry" list of all possible defects. This ordinarily is developed from engineering specifications and performance data, inspection records, field complaints, and the experience of inspectors, foremen, and engineers.

It is important to keep the end use of the item in mind, and to make sure that the inspection plan encompasses the requirements of the user. It is easy in the manufacturing area to think in terms of good workmanship and engineering specifications, and to overlook the user.

The second step is that of classifying the defects by degree of seriousness, and it is customary to use three or four categories. Most of the companies using this technique use the terms of Critical, Major, Minor, Incidental or Minor B.

What constitutes a Major defect for example, depends on the product, its use, and the conditions of sale. We have chosen to define a Major defect as:

1. Defects which for their proper correction requires fixtures, tools, or gages not carried by the customer engineer.

2. Defects which require the removal and disassembly of a major unit.

3. Defects which cannot be readily identified, ie: electrical connections.

This definition is completely different from the customary definitions, as most of our machines are leased, and the customer engineers service the equipment without charge to the customer. The seriousness of a defect must, therefore, be considered in terms of the loss of use by the customer, and the cost of service to the company. This point is made because each application of these techniques should be considered by itself, and the quality engineer should not be bound by the text book definitions, unless of course, he feels that they adequately describe his particular situation.

The proper classification is best decided at a small conference where field, assembly, engineering, and quality control management is represented. It is through the group thinking that the seriousness of various defects can be correctly evaluated. If these various departments take part in developing the plan, the first step toward getting their cooperation has been accomplished.

DEMERIT VALUES

After the classification has been established, it is customary to assign demerit values for the purpose of arriving at a single figure or index which can be interpreted as a measure of the quality of a unit.

Again, it is important in assigning demerit values to have the group thinking of those who are concerned with the quality of the product.

The demerit value is a factor which is applied to a defect to give it emphasis in relation to other defects, so that the sum or average of a number of defects has significance as a measure of quality.

Developing the relative position of critical and major defects is rather difficult. There are a number of standards that can be applied, but they will be different for various types of products. For example, in accounting machines we think more in terms of the service time required to analyze and correct the defect. Undoubtedly, in an aircraft plant they think more in terms of safety. Probably an appliance manufacturer will think more of the appearance, salability, and trouble-free operation.

It is rather interesting to note that of the nine "demerit weighting plans" which appear in Figure 1, that 7 of the nine have a ratio of critical to major of between 2 and 3 to 1.

In 6 of the nine plans, the ratio of major to minor is between 4 and 5 to 1.

USE OF DOUBLE CODE

In setting up all of the possible defects that can occur in a product, it is easy in a particularly complicated unit to run into a very long list, especially if an attempt is made to identify the trouble. For example, relay contact points can be out of adjustment, loose, burned, dirty, grounded, shorted, and a vacuum tube can have an open filament, a grounded element, a low emission, etc.

Usually, it is desirable for corrective action to identify a defect with sufficient exactness that the cause can be deduced. Rather than list six or seven possible defects that might be wrong with a particular element, we have established a double code, one that identifies the item that it appears will be required, and a trouble code which describes what

COMPARISON OF
DEFECT CLASSIFICATION AND
DEMERIT WEIGHTING

CLASS LETTER OR NUMBER	CLASS NAME	ELECTRONIC MFG. DIV. HUGHES AIRCRAFT CORP.	WRIGHT AERONAUTICAL DIV. WOODRIDGE, N.J.	G.E. & A.F.Q.C. PLAN LYNN, MASS.	JURAN, QUALITY CONTROL HANDBOOK	BELL TELEPHONE SYSTEM	HQ. 15TH A.F. MARCH A.F. BASE	H.A.C.Q.C. POLICY MANUAL	FEDERAL TELEPHONE SYSTEM	IBM CORP.
A I	CRITICAL	10	10	50	100	100	500	10	50	–
B II	MAJOR	5	5	25	75	50	100	5	20	15
C III	MINOR	1	1	10	25	10	5	1	5	5
D IV	INCIDENTAL	–	.1	5	5	1	1	–	1	1
E V	MINOR B	–	–	–	–	–	–	.1	–	–

FIG. 1

is wrong with the detail. A combination of the two makes it possible to identify the defect to a sufficient degree. The trouble code is shown in Figure 2.

TROUBLE CODES	
01	SHORTED/GROUNDED
02	STICKING/BINDING
03	STRIPPED
04	TIMING
05	WEAK/OPEN
06	WORN
07	SURFACE FINISH
08	DAMAGED
09	INCORRECT/MISSING
10	ADJUSTMENT
20	ALIGNMENT
30	BENT/WARPED
40	BROKEN
50	BLOWN/BURNED
60	DIRTY/CORRODED
70	DRY
80	LEAKING
90	LOOSE

FIG. 2

There are a number of points that generally come up in defect evaluation. One is the effect of "extent of deviation from specification." If an adjustment is only slightly outside of specification, its effect on performance may be minor, but if it deviates substantially from specification, it may become critical. It has been our practice to assign just one classification to a defect, and not attempt to make this refinement. It could be done in certain important mechanisms, and it will undoubtedly be considered as we make refinements in our plan.

Another point to be considered is the scoring of the number of defects. If an operator fails to lubricate a mechanism, for example, should every dry bearing be scored as a defect, or should the lack of proper lubrication be scored one defect? The same thing occurs when scoring any defect that is repeated because of single oversight or wrong starting adjustment. It makes a big difference how these items are scored as the overall quality index may be distorted if each individual defect is counted. The solution is how the defect is created, and what effect on product quality will be.

Another consideration is the establishment of a performance standard. If the organization is committed to improve present performance, there is

nothing to be gained in having a standard.

Some plans call for a "base period" standard. The average performance during a period which is regarded "normal" is taken as the standard. Others establish an "ideal" based on the number of possible defects in a unit. This would, for most products, be rather difficult to obtain. An approach might be made on the basis of recording the total number of different defects over a long period of time, and establishing the normal probability of occurrence of each individual defect. The total probability would then be the sum of the individual probabilities.

If a standard is necessary, a good start can be made by tabulating the defects for a period, looking at the distribution, and setting a standard on the basis of judgement.

ESTABLISHING AQL VALUE

If it is planned that the product is to be sampled, then the AQL value must be established by analysis of the requirements of the consumer. Setting this value will automatically setup a standard, and failure to meet it will force 100% examination of every item.

As a case in point, we were purchasing a rather complex unit from a subcontractor, and it had been necessary to do 100% inspection to maintain a proper quality level. Since it was not possible to sample the unit because the subcontractor had not been able to meet the standard, the burden of 100% inspection fell on us, and there was no incentive for the subcontractor to better his quality. The situation was solved by setting up two quality levels. If a sample indicated that the first level was met, it was agreed that we would 100% inspect to reach the second; if not, the material was returned to the subcontractor. The quality finally improved to the point that the second level was reached, and now the units are accepted on a sample basis. The "double standard" served its purpose of setting an immediately attainable quality level, and the returned lots provided the subcontractor with the incentive to improve his quality.

CONTROL LIMITS

It is possible to compute control limits for demerit charts in the same manner as they are computed for "P" charts. The formula is

$$\sigma_D = \sqrt{\frac{C_s}{n}}$$

$$\text{where } C_s = W_1^2 P_1 + W_2^2 P_2 + W_3^2 P_3$$

for a three-fold system of classification. W_1, W_2, and W_3 are the demerit weights assigned to each of the classes. P_1, P_2, and P_3 are the average or "expected" fraction defective. n is the sample size.

It is common practice to use 2 σ limits in the shop to start corrective action machinery into action when the chance is about 5% that there is no assignable cause, and 3 σ limits on reports to management where there is only about .25% chance that there is no assignable cause when a point exceeds the limit.

ANALYSIS OF QUALITY DATA

Next to controlling quality, probably the most important function to

be served by a formal technique is that of pinpointing the need for corrective action.

Recording quality data in punched cards not only simplifies the clerical task of computing the average demerits/per unit, it makes possible a detailed analyses which is the first step in directing the corrective action. A typical card of this type is shown in Figure 3.

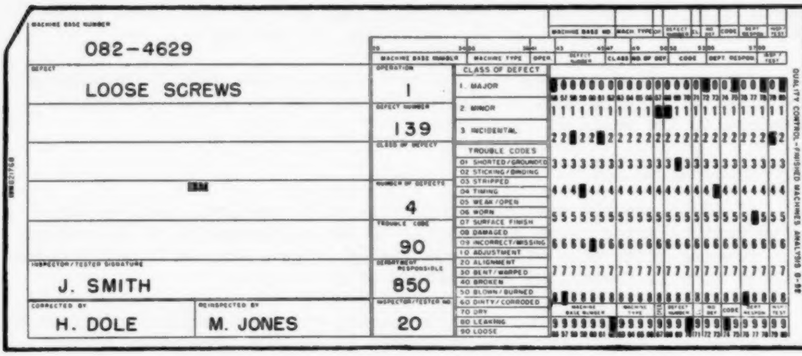


FIGURE 3.

By sorting the cards and tabulating them by defect number, it is possible to produce a report showing the frequency of occurrence of each defect. Scanning the report quickly indicates those items requiring attention.

The same cards can be regrouped and tabulated by department responsible. This makes it possible to produce a report to be sent to the foreman or manager responsible, itemizing each defect, the frequency of occurrence, and its contribution to the quality of the product. Such a report is shown in Figure 4.

The same cards can be regrouped by inspector and tabulated to show the comparative performance of a group of inspectors who are inspecting the same product. It can be very enlightening to review the difference in inspectors. It may indicate a need for more training of certain inspectors, and is a fine tool of supervision for the inspection manager. These are but three of many types of reports available to analyze quality data.

The accounting machine can handle a tremendous volume of quality data so rapidly that many quality analyses which would be prohibitive to do by manual methods are practical on a monthly, weekly, and even daily basis.

AUTOMOTION AND QUALITY CONTROL

Looking ahead, it is within the foreseeable future that we shall have additional and more elaborate testing robots that can be attached to a product that will record quality data in printed form, and automatically summarize, as well as analyze, mechanical, electrical, and electronic defects.

IBM

FINISHED MACHINES QUALITY ANALYSIS

1. BY DEFECT NO. [X]
2. SUMMARY BY UNIT AND DEFECT NO. [ ]
3. BY DEPARTMENT RESPONSIBLE [ ]

WEEKLY MONTHLY X X PERIOD ENDING DEC. 31 1953

CLASS OF DEFECT
1. MAJOR
2. MINOR
3. INCIDENTAL

TROUBLE CODES
01 SHORTED/GROUNDED
02 STICKING/BINDING
03 STRIPPED
04 TIMING
05 WEAK/OPEN
06 WORN
07 SURFACE FINISH
08 DAMAGED
09 INCORRECT/MISSING
10 ADJUSTMENT
20 ALIGNMENT
30 BENT/WARPED
40 BROKEN
50 BLOWN/BURNED
60 DIRTY/CORRODED
70 DRY
80 LEAKING
90 LOOSE

MACHINE TYPE	OPER. NO. OR DEPT. RES.	DEFECT NO.	DEFECT NAME	CLASS OF DEFECT	TROUBLE CODE	NUMBER OF DEFECTS
999	2	100	MOTOR & PUMP UNIT			
	2	101	BRUSHES	2	1	1
	2	101	BRUSHES	2	4	114
	2	101	BRUSHES	2	8	69
	2	101	BRUSHES	2	9	10
	2	101	BRUSHES	2	10	29
						223
	2	102	CAPACITOR	2	1	4
	2	102	CAPACITOR	2	5	11
	2	102	CAPACITOR	2	8	2
	2	102	CAPACITOR	2	10	3
	2	102	CAPACITOR	2	40	1
						21
	2	103	MAGNET ARMATURE	1	4	4
	2	103	MAGNET ARMATURE	1	10	4
						8
	2	105	CONTACT UNIT	2	4	1
	2	105	CONTACT UNIT	2	10	3
						4
	2	107	MAGNET LATCH	2	90	2
						2
	2	108	CONTACT UNIT	3	90	5
						5
	2	109	START CONTACT	1	10	1
	2	109	START CONTACT	1	90	1
						2
	2	114	WIRING	1	9	3
						3
	2	115	IMPELLOR BLADE	2	10	30
						30
	2	116	DEFLECTOR	2	2	2
						2
	2	117	GUIDE	2	7	48
	2	117	GUIDE	2	90	23
						71
	2	118	MP BASE UNIT	2	2	1
	2	118	MP BASE UNIT	2	4	8
	2	118	MP BASE UNIT	2	7	1
	2	118	MP BASE UNIT	2	10	5
	2	118	MP BASE UNIT	2	90	4
						19
						*390
			UNIT NAME AND FIGURES ARE FICTITIOUS			

CLASS I DEFECTS	CLASS II DEFECTS	CLASS III DEFECTS	NUMBER OF MACHINES	NUMBER OF DEFECTS
2594	3817	1718	3390	8129

FIGURE 4.

Already some robots of this type have been built, and it is only a matter of time until we shall all be using them.

PERFORMANCE TESTING

In order to obtain a relative measure of product quality through performance tests, the demerit technique can be applied.

We periodically select a finished production machine, and send it to the Testing Department where it is critically examined by engineers, and put into operation on a round-the-clock basis. As defects develop, they are evaluated and classified as Major, Minor, or Incidental and given a demerit value. While the test engineers do not have a formal classification of defects, the definition of each class is so objective that it is possible for each defect to be definitely classified at the time it is discovered.

By accelerating the test, aggravating the environment, and operating the unit continuously, it is possible to obtain the equivalent of normal life in a few months.

The test can be terminated at any time, but the most logical length of performance test is for the "expected" life of the product. There could be some justification for using the "guarantee" period, but ordinarily this is so short that the relative quality of units has little opportunity to be measured.

Unfortunately, the conditions of actual use are not duplicated, and the differences must be taken into account in evaluating the results.

At the end of the test, the vital units are disassembled and examined for evidence of abnormal wear and impending failure. Each condition observed is classified and given a demerit weighting, and it is customary in the preparation of the narrative summary which supports the demerit rating to assign the responsibility to manufacturing or engineering so that corrective action is started by the group thought to be primarily responsible.

In our plant, the cost of the unit and length of test necessary requires that the number of machines to be tested be carefully weighed. Undoubtedly, companies manufacturing a less complicated and less expensive item in larger quantities would find it practical to test many more.

The scoring system used by our Testing Department is essentially the same as that used in Final Inspection --

Major	13 to 17 demerits
Minor	4 to 7 demerits
Incidental	1 to 2 demerits

A range is permitted the test engineer in scoring a defect as it is felt that his training, experience, and the thoroughness of his examination make it possible to weigh each defect with greater discrimination.

The Testing Department has defined Major, Minor, and Incidental defects as follows:

MAJOR:

1. Defects which, for their correction, require gages, fixtures, or tools not carried in the domestic tool kit.

2. Defects which require removal or disassembly of a unit for their correction. Units, such as counters, which are designed for easy removal are not included.

3. Poor electrical connections or contacts which cannot be identified as such by visual means.

4. Safety hazards which require exercise of unusual precaution.

5. Troubles which require replacement of parts not ordinarily carried by the Customer Engineer.

6. Conditions which, if neglected, could result in defects as outlined in "1", "2", "4", or "5" of the major classification within a 90 day period.

MINOR:

1. Defects which require corrective action and can be readily accomplished by means of customer engineering tools without disassembly or removal of a major unit.

2. Potential safety hazards. Possibility of being cut or pinched. (Common sense and a little caution might prevent an accident.)

3. Defects which, if neglected, could result in defects as outlined in "1" of the Minor classification within a 90 day period.

INCIDENTAL:

1. Defects which do not affect the operation of the machine, and which, if neglected, would produce a machine malfunction.

2. Small imperfections in machine appearance, not serious enough to produce customer dissatisfaction.

In the course of time, we hope to be able to find a relationship between the test result and final inspection and field performance data. To date, there does not appear to be any relationship, but a detailed analysis of the defects reported and their impact on the final score or quality index over a period of time should explain the reasons for the difference.

FIELD PERFORMANCE

The measurement of quality in terms of field performance also involves selecting an appropriate index as representative of quality in general. There are many. We noted in an article that appeared recently in one of the technical publications that one company used the number of returned units as an index. The number of service calls during the guarantee period, the dollar value of replacement parts, the number of complaints, etc., have been used to indicate the quality of the manufactured product. There is probably no one indicator that might be universally used.

As Professor Juran points out in his Handbook, the number of complaints may depend too much on the condition of the market, the customer's inventory, his temperament, and other factors not related to the quality of the product.

Due to the fact that many of our products are leased and the service is part of the rental, we have chosen the average service hours per unit per month as the quality index. This index was selected because the time spent servicing the equipment was felt to be related to the quality in such a way that the more serious a defect, the longer it ordinarily takes to discover and correct it, and the hours spent correcting defects was an obvious indication of the cost of service, and in management's thinking certainly related to the quality built into the product. However, we have discovered that it is not the perfect index of quality.

For example, time required to service a unit is obviously related to the customer engineer's ability to diagnose and make the proper repair. When a machine is a new model, it takes longer than when it is more familiar to the customer engineers. A customer engineer in a provincial territory covers a greater variety of machines, and it is likely to take him longer to diagnose and repair a trouble on a particular machine than it would a specialist in the city.

In order to balance and randomize the many variables likely to become involved in this field performance evaluation, we consulted an expert on polling and market surveys. He studied the problem, and on the basis of geographic area, type of application, type of service required by the customer, and the nature of the territories (urban vs. provincial), selected a group of offices to be surveyed.

SELECTION OF PERIOD

The selection of the proper period also posed a problem. There were some managers and quality engineers involved who felt that the defects found at the time of the installation and test should be used, others who felt that the performance in the first 90 days should be taken, and others who felt that the first year should be used.

I believe that a lot depends on the type of product, the type of inspection, and test given the product at installation. If the product is thoroughly examined and/or defects are likely to appear at installation, it is conceivable that the defects found at installation would be a good index of product quality.

With a complex unit and many parts, it is likely that troubles will develop which could not be discovered at the time of installation. It can be logically argued also that these latent troubles can remain hidden for longer than 90 days. For example, defective heat treating resulting in soft parts may not show up for six months or even a year.

The pros and cons of this were debated and finally it was agreed that for the purposes we had in mind, that any field troubles which appeared in the first 90 days of operation would logically be the responsibility of the manufacturing organization.

The longer the period is extended, the more the quality of the design begins to enter into the index. A part or unit which is structurally inadequate to stand the rigors of prolonged service may have been

manufactured properly and have failed because the design is not sound.

So we ended up by preparing a quarterly report of the 90 day performance of machines shipped to the sample offices, and what we have chosen to call an Age Study by Year of Manufacture. This is a summary of service experience on all machines that are 1, 2, 3, 4, and 5 years old. This latter report is used mostly by Engineering to evaluate the soundness of design, the "expected" life, the effect of improvements in design, and other items of interest.

A sample of the 90 Day Trouble Analysis is shown in Figure 5. This summary sheet shows the distribution of calls and service hours by unit. The detailed listing by unit, part number, and trouble code appears as part of the report.

The Age Study by Year of Manufacture is prepared on the same form.

For comparative purposes, the average service hours per unit per month is plotted on a chart which is distributed together with the charts that show the result of product test and final inspection. This chart is shown in Figure 6.

It is the function of the Quality Control Division to gather all this data, analyze it to the point that the areas of responsibility are fairly well indicated, and publish the quality data for the use of assembly and manufacturing managers.

In our company, quality has always been the responsibility of the operating organization, and corrective action is initiated by the responsible manager. The quality control organization assists by making studies, evaluating quality reports, and inspecting the product to see that the quality is there.

ANALYSIS OF QUALITY REPORTS

Since quality reports concern themselves with measuring an intangible, it is important that managers or foremen be trained to use them properly.

There is a natural tendency on the part of the manager or foreman to be critical of a report that shows him to be responsible for certain defects. Instead of looking for the items that he can do something about, he concentrates on finding defects charged to him that belong to someone else. The report's effectiveness can be spoiled if the emphasis is allowed to center on "who" is responsible, rather than on "what" is responsible.

When attention is focussed on some particular condition because of its frequency of occurrence or seriousness, it is important to have sufficient detail that the defect is clear, and the individual responsible knows what must be corrected. In the case of final inspection and field performance report analysis, we obtain the actual call report cards and group them so that when any particular trouble is to be investigated, they can be reviewed to completely clarify the condition. The narrative account that the customer engineer makes on his call report can be very informative to one who is charged with the task of finding the solution to reduce the recurrences of a particular trouble.

IBM

FORM TO-0000-0

EAM

90-DAY TROUBLE ANALYSIS

QUARTER 4TH MONTH___ WEEK ENDING ___ 53 CUSTOMER ENGINEERING DEPARTMENT

MACHINE TYPE	UNIT CODE	PART CODE	UNIT OR PART NAME	TROUBLE CODE	MACHINES ANALYZED: QUANTITY BY TYPE	TOTAL FOR PERIOD: CALLS	TOTAL FOR PERIOD: HOURS	ALL MACHINES: ANNUAL HOURS *	AVER. PER MACHINE PER MONTH: CALLS	AVER. PER MACHINE PER MONTH: HOURS	% OF UNIT OR PART HOURS TO TYPE HOURS	RANK WITHIN TYPE: UNIT	RANK WITHIN TYPE: PART
999	80		CONTACT UNIT		94	255	10151	4060	9	36	238	1	
999	06		RELAY UNIT		94	209	6114	2446	7	22	143	2	
999	10		FEED UNIT		94	171	3483	1393	6	12	82	3	
999	00		BASE UNIT		94	131	3232	1293	5	12	76	4	
999	04		PUMP UNIT		94	68	2806	1122	2	10	66	5	
999	50		SWITCH CONTROL UNIT		94	92	2407	963	3	9	56	6	
999	60		GEN RECT UNIT POWER SUPPLY		94	44	2103	841	2	7	49	7	
999	09		MOTOR AND DRIVE UNIT		94	45	1335	534	2	5	31	8	
999	48		MAGNET UNIT		94	10	994	398		4	23	9	
999	02		MOTOR CLUTCH UNIT		94	16	969	388	1	3	23	10	
999	05		TUBE CONTROL UNIT		94	25	861	344	1	3	20	11	
999	40		IMPELLOR UNIT		94	29	507	203	1	2	12	12	
999	30		AUTO VALVE UNIT		94	40	495	198	1	2	12	13	
999	12		CLUTCH UNIT		94	11	219	88		1	5	14	
999	03		EXHAUST UNIT		94	10	73	29			2	15	
999	42		CAM UNIT		94	1	45	18			1	16	
999			NO UNIT MENTIONED		94	133	3673	1469	5	13	86		
999			NO TROUBLE FOUND		94	35	2066	826	1	7	48		
999			CUSTOMER ERROR		94	52	894	358	2	3	21		
999			CUSTOMER POWER SUPPLY		94	5	179	72		1	4		
999			UNASSIGNABLE UNIT NTF		94	2	25	10			1		
999			CUSTOMER ATTACHMENTS		94	1	102	41			2		
			TYPE 999 MACHINE		94	1385	42733	17093	49	152			
999			ASSIST TIME		94		5482	2193		20	128		
			PERIOD SAMPLED 13 WEEKS										
			UNIT NAMES AND FIGURES ARE FICTITIOUS										

FIGURE 5
****

EAM TROUBLE CODES

01 - Adjustment
04 - Alignment
07 - Bent-Warped
10 - Blown
13 - Broken
16 - Burned
19 - Card Jam
22 - Dirty-Corroded
25 - Dry
28 - Leaking
31 - Loose
34 - Off Punched Cards
37 - Open
40 - Shorted
43 - Sticking-Binding
46 - Stripped
49 - Timing
52 - Weak
55 - Worn
[illegible] - Other
[illegible] - No Trouble Found
[illegible] - Incomplete (Applies to part encoded groups only)

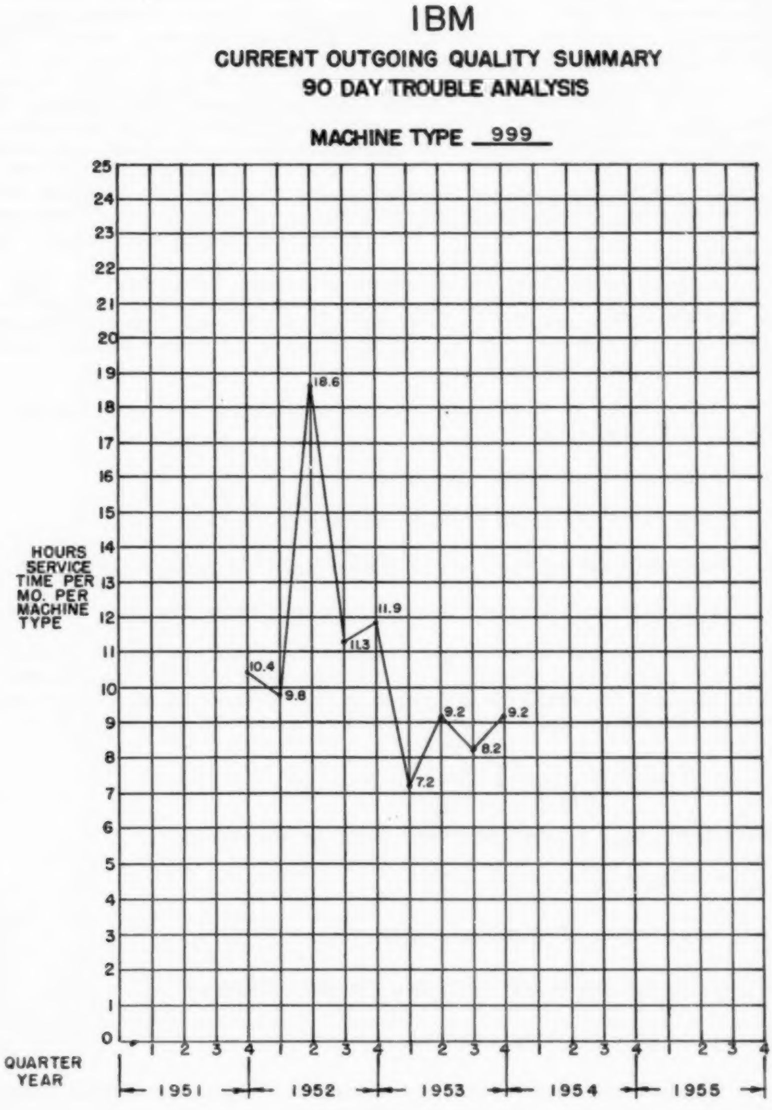

IBM
CURRENT OUTGOING QUALITY SUMMARY
90 DAY TROUBLE ANALYSIS
MACHINE TYPE 999
HOURS SERVICE TIME PER MO. PER MACHINE TYPE
QUARTER
YEAR
1951
1952
1953
1954
1955
10.4
9.8
18.6
11.3
11.9
7.2
9.2
8.2
9.2

FIGURE 6
****

To be effective, reports must be timely. To make the operating management completely happy, we need the quality picture the day after the product is made. It seems that the older a report becomes, the less attention it receives, and a report on products that were shipped three or four months ago is of value only in studying trends.

If the report is on the 90 day field performance of a product, it obviously must represent products that were manufactured at least three months ago. Therefore, it is not always possible to be timely in reporting quality. It is, however, a very important consideration, and any quality evaluation planners should consider it one of the foremost items in establishing their system of reporting.

COMBINATION OF QUALITY INDEXES

It has been suggested that a group index representing a combination of all products manufactured in a given plant is needed to give the top executive a figure to guide him in maintaining the proper balance between production, cost, and quality.

To do this, it is necessary to use a weighting factor or factors. Among those used are:

1. Dollar value of various products.

2. Total number of units produced of various products.

3. Complexity factor based on estimated number of possible defects.

4. "Expected" or "Standard" demerits per unit value for each component.

It is possible to develop a standard deviation and control limits for this group index of quality in the same manner as for the demerit per unit charts. Dr. Romig in his paper on the "Evaluation of Quality Through Demerit Rating System", gives the standard deviation of the index as

$$\sigma = \frac{1}{\sum D_{u_{s_i}}} \sqrt{\sum \frac{C_{s_i}}{n_1}}$$

Where the $\sum D_{u_{s_i}}$ = the "expected" demerits per unit

$C_{s_i} = W_1^2 P_1 + W_2^2 P_2 + \text{---} W_n^2 P_n$

n_1 = total number of units in sample

W_1, W_2, -- W_n = the weighting factor for each component

P_1, P_2, -- P_n = fraction defective

We might add that we have not chosen to develop any overall indexes of quality, because it is generally felt that it is necessary to develop confidence in the individual measures before attempting to combine them.

The first year that measures of this type are used to evaluate quality can well be spent reviewing the results, and making improvements in the system. It is possible that as our confidence in the system develops,

that we shall undertake to develop an overall index.

CONCLUSION

It has been truly said that "to control, you must be able to measure." These measures are the most effective controls developed to date. Back in 1928, Mr. Harold Dodge, in an article published by the Bell Telephone Laboratories entitled "A Method Rating the Manufactured Product", outlined the principle features of the demerit rating system.

It has grown in application because it is fundamentally sound, and is presently being applied in hundreds of plants to measure the quality of all types of manufactured products.

There is much to be learned, however, in the field of application of this technique, and the purpose of this paper has been to review some of the practical problems that crop up when it is applied.

TEXTILE QUALITY ANALYSIS

Charles C. Wilson
Research Division
West Point Manufacturing Company

The textile industry has a very interesting history. It is one of the oldest factory industries and, of course, the history of textiles in the home extends back to the dawn of civilization. We identify the first textile factories with the beginning of the industrial revolution in the eighteenth century and associate them with the invention of the ring spinning frame, the power loom, and the cotton gin. In contrast, many industries have developed from more recent scientific discoveries and advanced technical knowledge. These modern industries have grown up with scientific research and laboratory testing as an integral part of their organization.

The textile industry, on the other hand, is a remarkable example of technological development. Textile technology has developed from the experiences of generations of capable, energetic men whose ingenuity and resourcefulness have been sharpened by competition and the ever critical demands of the consumer. Skills, machines, and ideas developed by the patient and persistent efforts of countless individuals and groups, are our common heritage. This accumulation of inventions and refinement of technical knowledge has steadily improved quality, increased production, and diversified textile products. The immensity of this heritage, its thoroughness and scope is particularly impressive when a search of patents is made on any phase of textiles.

There comes a time, however, when technological progress slows down or even stops because of incomplete understanding of specific facts or for the lack of certain scientific discoveries. Research and scientific investigations are organized efforts to hasten this process and promote technological development in an efficient manner. This general pattern is especially true for textiles. Textile technology was reaching a slowing-down period when a series of events sparked by the development of chemical fibers reversed the trend. Currently, textile technology is experiencing an accelerated rate of progress.

As might be expected, the first years of statistical quality control in a textile mill have been difficult. Rule-of-thumb methods for control and authoritative opinion are well established. The qualities measured were those emphasized by operating management, and the test methods were those that had been handed down for generations with only slight modifications. Nevertheless, the new quality control methods proved successful, and some very important improvements have resulted. Quality control charts have been effective in revealing the lack of standardization of quality (1) between machines, (2) between operators, (3) from job to job, (4) from room to room, and (5) from day to day. After surviving this initiation, we have begun to take another look at textile quality. With the advantages of research investigations and some new testing instruments, we are changing some of our opinions about textile quality.

In this respect, industrial research and quality control have a common foundation. This foundation is quality analysis. Quality analysis is an effort to know what constitutes quality in the product, and understand the relative importance of the many factors that affect or cause quality differences. Without this information, the labor of

both research and quality control may be misdirected. In certain cases, a new testing instrument or testing method must be developed in order to make a complete analysis. This is often the first step that leads to a chain of reactions and further improvements in quality. In other cases, industrial processes, like magicians, are quite successful in deceiving us about what actually is happening. Our attention is sometimes too firmly focused on the more obvious variables, and we fail to see the effects caused by others.

Let us first review some of the changes that are taking place in textile fibers. For hundreds of years, cotton, wool, and other natural fibers have been bought and sold on experts' judgment of spinning quality. Grade and staple classification is the basis for cotton marketing, although other factors are taken into consideration in the purchase of <u>spot</u> cotton. The mill buyer considers also in a general way the area of growth, varietal characteristics, the effects of weather, insect damage, or any other factor that may affect quality. In a somewhat similar manner, the raw stock quality of all natural fibers has been controlled by authoritative opinion.

During the last quarter of a century, much additional information has been learned about textile fibers by laboratory testing and experimental investigations. Development of a variety of chemical fibers with distinct physical and chemical characteristics has accelerated these studies. From these raw stock quality analyses, we have a much better understanding of what determines spinning quality. Practical application of this knowledge to quality control of raw stock has been retarded. Most laboratory tests for fiber quality are too slow for extensive use in cotton buying. A mill's supply of raw cotton may come from thousands of farms scattered over several states and representing the effects of extensive environmental effects. Traditional marketing systems tend to ignore these differences and as a result, sample inspection of incoming lots is not entirely satisfactory. Laboratory tests can be used to some advantage in selective blending in the mill, however.

The Micronaire for measuring fiber fineness is an exception. This test is rapid enough for 100% inspection of cotton bales. American Upland cotton ranges from 2.5 to 6.5 micrograms per inch, and the range in fiber fineness for the world supply is somewhat greater. Although cotton classers have some ability to distinguish degree of fiber fineness, little or no attempt was made to use this skill. This is rather surprising because the commercial grading of wool is based upon fiber fineness, and denier of man-made fibers is recognized as an important fiber characteristic. Also, the importance of cotton fiber fineness was brought to our attention at least twenty years ago. In the last five years, control of fiber fineness by Micronaire testing has become a routine quality control procedure in most mills. This method measures the rate of air flow through a standard weight sample confined in a special container. The test requires about two minutes. Previous measurements of fiber fineness required about two hours testing time. Since it has become practical for the mills to control this fiber quality, important improvements have been experienced in running performance, evenness in dyeing, reduction in neps, and improvement in finishing.

This is only one case where a better understanding of what constitutes fiber quality, plus a new scientific means for measuring it, has made very important changes in the textile industry in recent years.

Now cotton can be bought on a fiber fineness specification in addition to grade and staple. These are definite steps along the route of technical progress. Let us not forget, however, that there are other fiber qualities that determine spinning value. The processing performance and yarn quality are also affected by other fiber qualities: (1) length and length uniformity, (2) strength and elongation, (3) shape or configuration, (4) maturity, (5) surface finish, and (6) physical and chemical degradation due to exposure, mechanical treatment and attack of microorganisms. In recent years, a large amount of experimental investigation has been directed toward measuring these variables and associating them with spinning quality. These quality analyses will eventually bear fruit. When the laboratories provide us with faster measuring instruments and commercial marketing practices are altered to meet the needs, control of these variables will then become technical achievements.

Quality analyses are also being applied to manufacturing processes within the mills. Experimental investigations employing improved testing equipment and statistical analysis are examining some of the intangible relationships between processing and textile quality. Conclusions from these analyses don't always agree with established opinions. The industry moves cautiously. New evidence must be completely digested and assimilated in its massive bulk. A clear understanding of the relationships is important. That is the function of quality analysis. Total variation in quality is examined in terms of the sources that produced it. Each source is evaluated according to how much it contributes to total variation. Systematic and cause variation must also be clearly distinguished from random or chance variation.

An example of quality analysis of a textile process can be illustrated by a study of picker lap quality. Picker laps are the first of a series of intermediate products between the raw stock and the finished product (Fig. 1). A typical cotton lap may have these specifications: 40" wide, 55 yards long, 14.5 ounces per linear yard, and a total lap weight of 50 pounds. Quality control at the mill has been directed toward controlling those qualities that are most easily measured, namely total lap weight, and yard-to-yard variation. Charts comparing lap yard-to-yard variation in two laps are shown in Fig. 2. The laps compared represent the product of two different pickers from different mills. By mill standards, one lap would be considered satisfactory and the other of very poor quality. An inquiring mind is not completely satisfied and insists upon the analysis of quality in terms of its effects upon subsequent quality.

Lap stock is fed to carding machines which in turn produce a rope-like strand called sliver. Card sliver weight per yard is about 60 grains, resulting from a carding draft of 100 or more. One-third inch of lap stock is equivalent in weight to about one yard of card sliver. Short length variation in the picker lap can, therefore, result in card sliver weight variation also. A lap tester was developed to measure short length variation in picker laps. Fig. 3 shows sections of charts from another set of laps from the same two pickers. Lap quality compared on this testing machine ranks the two pickers in reverse order. The picker that produced laps very poor in yard-to-yard uniformity was considerably better in short length uniformity. Fig. 4 takes quality analysis one step further.

Laps from these two pickers were carded on the same carding machine

and all of the sliver from each lap tested for sliver weight per yard uniformity. Every tenth yard was weighed and recorded in consecutive order. The graphs on the left represent an analysis of picker lap irregularity, reconstructed from the card sliver weight per yard data. The heavy center line is the average sliver weight per yard obtained from each consecutive yard of picker lap. The upper line represents the heaviest individual yard of card sliver from each consecutive yard of picker lap. The lower line represents the lightest individual yard of card sliver from each consecutive yard of picker lap. The average range of sliver weight within each yard of picker lap is about 8.5 grains per yard for Mill A and about 6.5 grains per yard for Mill B. The frequency distributions on the right show total variation in card sliver weight per yard. The C. V.'s were 5.52% and 5.75%, respectively. Fig. 5 demonstrates the mathematical analysis of these data. From this analysis, we must draw the conclusion that lap short length irregularity is caused by a different set of factors and is produced independently of lap yard-to-yard variation. Both sets of factors are important to card sliver uniformity and should be controlled. This fact is not generally understood.

Not content with this laboratory test alone, we wanted to know what was actually happening in our plants. First, we recorded the weight of every lap doffed from a battery of pickers during a day. Second, we tested laps from each picker for yard-to-yard uniformity. Third, we measured short length fluctuation in the lap stock. These tests gave us information about our lap quality. Fourth, we determined the total amount of variation in card sliver weight per yard by taking specimens from all the cards every day for a number of days. By analyzing these data, we learned not only how much variation in sliver weight per yard existed in each mill, but also what produced it and the relative importance of each cause factor.

Fig. 6 presents data from two mills that contrast in certain details. The same data can be presented in the form of an algebraic equation, but the graphic presentation seems to be more effective. Square #1 represents lap weight variation. Square #2 represents lap yard-to-yard weight variation. Square #3 is the total variation in yard-to-yard lap weight resulting from differences in lap weights among total lap production. Square #4 represents only the short length variation that occurs within each yard of lap stock. Square #5 is the resulting short length variation found in total laps produced. Square #7 represents the total variation in card sliver weight per yard. Squares #1, #2, #4, and #7 were constructed to scale and represent actual mill stock variation determined from extensive measurements under normal operating conditions. Squares #3, #5, and #6 represent the resulting variation from two other sources, and their size was determined by calculations. Total variation in card sliver weight per yard (Square #7) was not completely accounted for by the actual measured variation in picker lap stock (Square #5). Square #6 was, therefore, necessary to represent all other sources of variation that contribute to card sliver weight variation. These might include (1) variation in moisture, (2) slight differences between cards and card settings, and (3)operating variables.

Analysis of picker lap quality was presented in detail to illustrate what can and is being done at each manufacturing process. The total variation in stock size delivered by a process must be accounted for. The total variation delivered can be explained by (1) short term

variation within each machine, (2) differences among machines, and (3) long term variation caused by changes in moisture content, channeling, and other causes preceding the process. By the proper experimental procedure and statistical analysis, the contribution of each of these sources of variation can be examined and appraised.

Short term irregularity in yarns and the preceding intermediate products have always been a serious textile problem and an important quality characteristic. An instrument for measuring short term uniformity of sliver and roving has been available since 1925. Instruments especially designed for measuring the evenness of yarn have been developed only recently. Before evenness testers were developed, boss carders and spinners depended upon visual inspection for judging this quality, and set and regulated their machines accordingly. Machine manufacturers had poor guidance in improving textile machinery, and the incentive was uncertain. Lacking a precise means for measuring yarn, roving, sliver, and lap unevenness, a great deal of tolerance was allowed and technological progress was slow. Now that we have instruments for measuring this quality, the factors that cause unevenness are receiving a great deal of attention.

Machines delivering uneven stock can be detected, and the cause found and corrected. Routine testing data furnish reliable information for establishing maintenance schedules. Evenness tests are reliable guides for setting rolls, proportioning draft, and other operational adjustments. They are particularly useful in locating faults that produce cyclic variation or periodicity in the stock. The long-range benefits from evenness testing will probably be very important. Machine manufacturers now have an instrument to guide them in improving the design and operation of the functional parts. Probably equally important, the spinner can better judge the operational differences between machines of different manufacturers and designs.

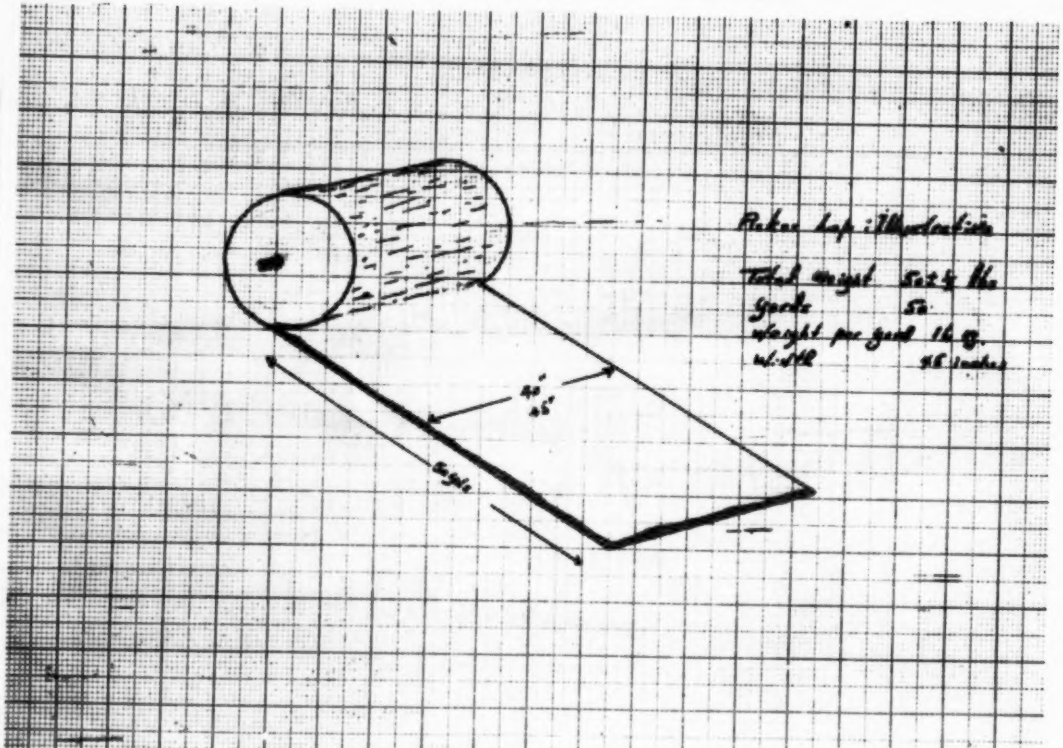
Quality analysis of unevenness is needed to guide these various efforts. We have some incomplete answers to a number of questions. Part of the short length unevenness in the product from one process becomes a longer length unevenness in the subsequent process because of drafting. Consequently, we have unevenness in yarn representing the sum total of the unevenness in all previous processes. Without a great deal of imagination, one can conceive that there might be a multitude of patterns. Some of these patterns are easily detected and their effects well known. For instance, a short term cyclic pattern in drawing sliver may become a glaring defect fillingwise in the fabric. Other unevenness patterns and their effect upon yarn and fabric quality are not so obvious. How do these various patterns affect processing performance, yarn strength, fabric appearance, fabric strength, air and water permeability, resistance to tear and abrasion?

Textile quality analysis must also consider the consumer's specific needs and demands. The end use may require a particular textile fiber or an appropriate blend of fibers. The fabric style and construction determine its appearance and qualities. In industrial fabrics we are especially interested in tensile strength, tear strength, resistance to abrasions, bursting strength, air and water permeability, and other special qualities. Many processing variables affect each of these qualities and in different ways. The right combinations are most important.

Textiles have been and will always be a major consumer product with

extensive and diversified uses. Technological knowledge concerning what fiber best suits each end use, and how to process, spin, twist and weave it into a wide range of fabrics to obtain these diversified qualities, is indeed a great heritage. Current demands for textile quality, however, are more exacting and diversified than at any time in history. There are many new fibers to help meet these demands. Competition between producers is sharp, and the consumers are critical. Authoritative opinions of experts and rule-of-thumb methods are inadequate for these demands. Operating on a balanced average is not enough. Intangible qualities of the raw stock are useful, but they are most valuable when understood and used properly. Yarn and fabric quality is affected by every process of manufacturing. Precise measurement of processing variables and analysis of quality is necessary for alert management. All signs indicate that quality control by laboratory testing will play a much greater part in textiles of the future.

Figure 1



Picker lap illustration
Total weight
Yards 50
weight per yard 16 oz.
width 45 inches
40"
45"
Single

Figure 2

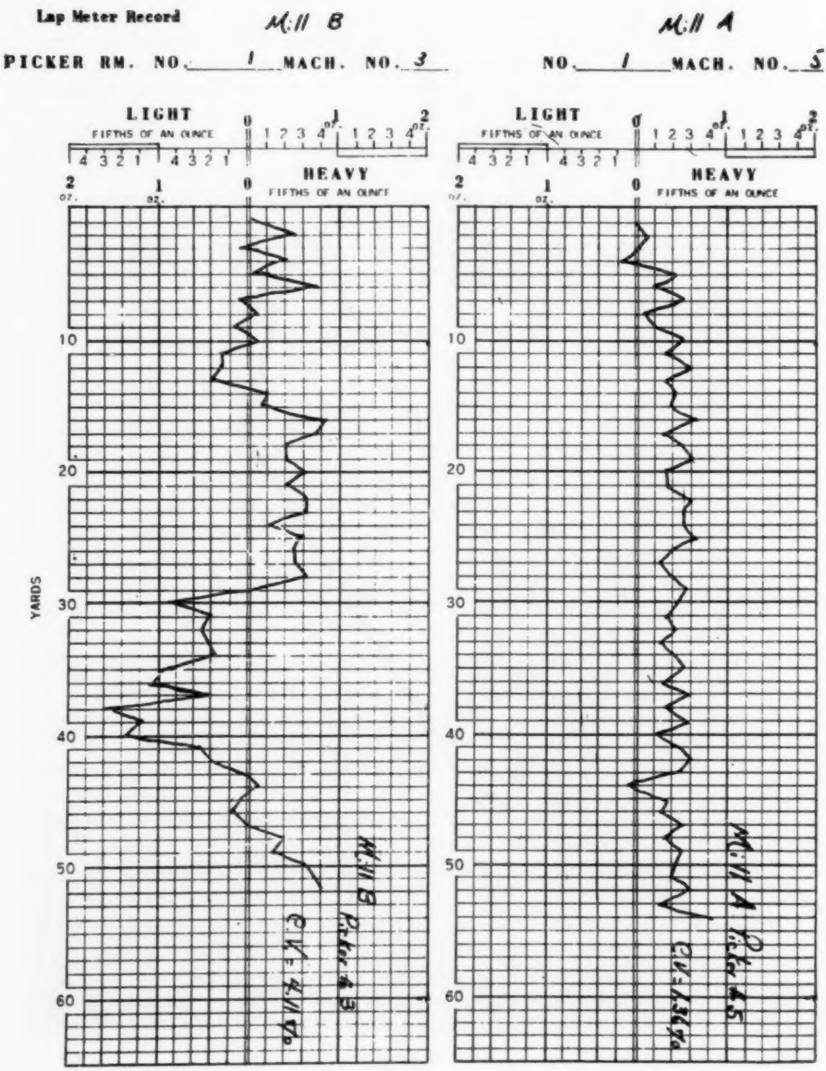

Lap Meter Record
Mill B
Mill A
PICKER RM. NO. 1 MACH. NO. 3
NO. 1 MACH. NO. 5
LIGHT
FIFTHS OF AN OUNCE
HEAVY
FIFTHS OF AN OUNCE
YARDS
Mill B Picker # 3
C.V. = 4.11%
Mill A Picker # 5
C.V. = 1.36%

Figure 3

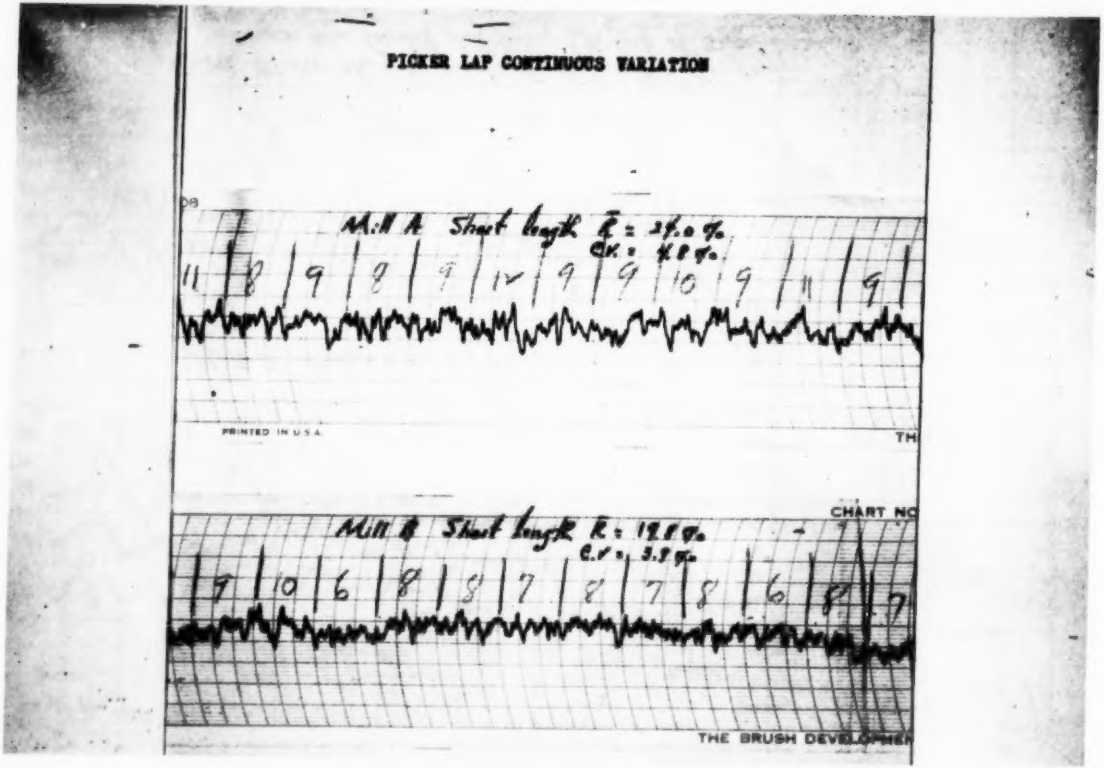

PICKER LAP CONTINUOUS VARIATION
Mill A Short length
PRINTED IN U.S.A.
Mill B Short length
CHART NO
THE BRUSH DEVELOPMEN

Figure 4

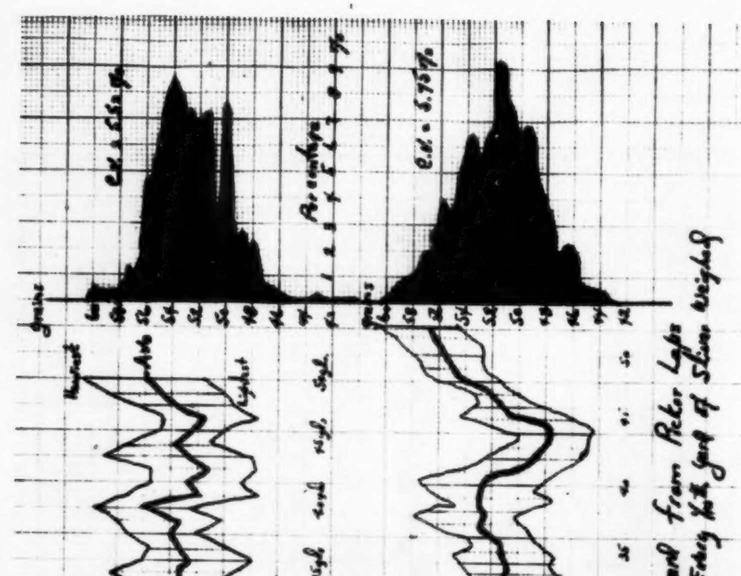

Mill A Picker #5
Mill B Picker #3
Card Sliver Weight per yard from Picker Laps

Figure 5

Ave. C.V. Yard - Yard (S.-L. Lap Meter) = 1.36%

Total C.V. Weight Per Yard Card Sliver = 5.52%

$V_T = V\omega + Vb$

$(5.52)^2 = V\omega + (1.36)^2$

1st Lap $\quad 31.00 = V\omega + 1.84$

$V\omega = 28.36$

$\sigma\omega = 5.42\% \times 52.5$ grains = 2.84 grains

$\bar{R} = d_2\omega = 3.078 \times 2.84 = 8.75$ grains

Ave. C.V. Yard - Yard (S.-L. Lap) = 4.11%

Total C.V. Weight Per Yard Card Sliver = 5.75%

$V_T = V\omega + Vb$

$(5.75)^2 = V\omega + (4.11)^2$

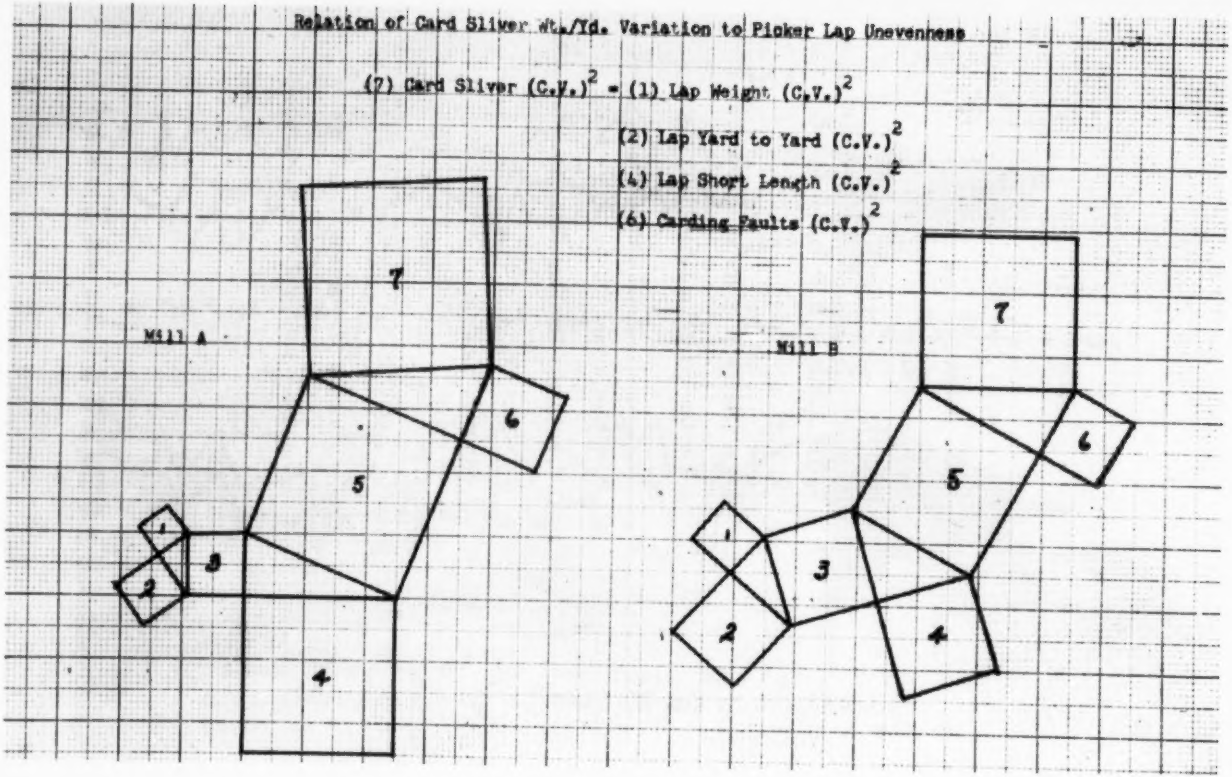
2nd Lap $\quad 3.30 = V\omega + 16.9$

$V\omega = 16.1$

$\sigma\omega = 4.0\% \times 52.0$ grains = 2.08 grains

$\bar{R} = d_2\omega = 3.078 \times 2.08 = 6.4$ grains

Figure 6



Relation of Card Sliver Wt./Yd. Variation to Picker Lap Unevenness
(7) Card Sliver (C.V.)^2 = (1) Lap Weight (C.V.)^2
(2) Lap Yard to Yard (C.V.)^2
(4) Lap Short Length (C.V.)^2
(6) Carding Faults (C.V.)^2
Mill A
Mill B
1
2
3
4
5
6
7

FOUNDRY QUALITY CONTROL FOR SHORT RUNS

Allin P. Deacon
Cockshutt Farm Equipment Limited

Introduction

Cockshutt Farm Equipment Limited is a Canadian Company employing about 5000 men and women among 5 plants, 4 in Ontario, and 1 in Ohio at Bellvue, under normal conditions. To-day employment in the farm equipment industry is temporarily low in Canada and the United States. The main factory, normally employing about 3000 people, is in Brantford, Ontario, where equipment is manufactured ranging in size all the way from small feed grinders to self-propelled harvesters which can cut and thresh a 15-foot swath of grain, and 50 horsepower tractors which can pull a 5-furrow plow through heavy soil.

Brantford is a city of about 50,000 people situated north of Lake Erie about 60 miles west of Buffalo. It is here that the mechanized foundry is located, producing castings for most of the implements manufactured by the company in Canada.

The variety of castings is considerable while the length of run on each casting is comparitively small. It is for this reason that we feel that the methods employed in improving the quality of our castings will be interesting to both large and small foundries.

The largest casting weighs 510 lbs. It takes 5000 of the smallest to fill an ordinary steel tote box 29 in. x 45 in. x 18 in. deep. Production runs vary from 25 pieces on repair orders to 25,000 on some small parts made for our regular run of seed drills. Most jobs call for regular gray iron that requires no special physical properties. Some large castings must be produced of iron with a tensile strength of 29,000 to 30,000 lbs. per sq. in. The plow shares must be produced of gray iron in the body of the casting with a white iron chilled surface along one edge controlled in depth to about $\frac{1}{4}$ in.

To meet this variety of requirements a base iron is produced in the cupolas which is inoculated in the ladle as necessary, using ferro-silicon or ferro-chrome as the particular jobs demand. The sand mix which is distributed to all the moulders' hoppers is designed to produce both the 510-lb. casting and the small one with 20 on a sprue.

The foundry is laid out as indicated in figure 1, with four cupolas and two sand mullers. The foundry is mechanized to the extent that the sand is distributed by conveyor belt from the two mullers to the 46 hoppers, the moulds are jolted and squeezed by machine, the moulds are transported on dollies to the pouring floor, the iron is transported in ladles hung on monorail trolleys, the dumped moulds of smaller castings are conveyed by belt below the floor level to a main shake-out, whence the castings go to the separating belt and the sand is conveyed back to the screener and storage hoppers above the mullers.

The core room is mechanized only to the extent that the sand is mixed in the muller above the core blowers and benches and distributed by belt to seven hoppers, and the raw sand is brought up to the muller by screw conveyors and weighing bucket.

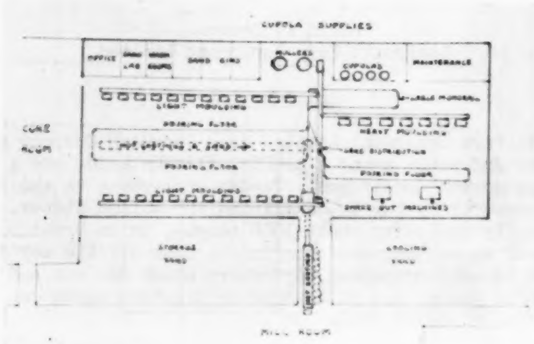


Figure 1

The mill or cleaning room is mechanized only to the extent of the automatic shot blast equipment.

Thus the set up requires methods of control used in fully mechanized foundries, but it also requires techniques which we believe can be used in smaller foundries without equipment such as ours and where production is not sufficient to warrant any large outlay for quality control.

History of Quality Control Program

The Approach

As in most foundries, all our castings go through a sorting process in the mill room, with all finished castings being subject to inspection by the inspection department before release to the stores department. All scrap is collected regularly during each shift and delivered to the scrap area for cataloguing according to the type of defect. All scrap castings are weighed before return to the foundry for remelting. The list of scrap is used by foundry supervision to trace sources of scrap. The weight records provide a measure of the scrap rate.

But our problem was to prevent scrap from being made at all.

In the initial stages of our program we had the advantage of having the full support of management, and, what was more important, that of the superintendent of the foundry. Our chief inspector, as head of our quality control department, was a keen supporter during the complete operation.

Our primary concern was to engage the interest and support of the union which is strongly organized in our shop. Once our plan of action was made, the union committee representative in the foundry sat down with us and a brief clear outline of our purpose and method was presented to him. His fair reception and discussion of our ideas and his agreement to a trial operation smoothed the way for our initial set up to be tried out

by ourselves. It developed later that he joined our staff as analyst of casting defects. He was a good moulder and he turned out to be a conscientious analyst.

Detection of Defects

A start was made by collecting samples of castings made by one operator using the multiple sampling plan from the MIL Standard 105A tables for 4% average quality level and an average lot size of 200 pieces. This involved collecting 10 pieces from the hot shake-out and letting them cool, then having them cleaned in the shot blast unit ready for inspection. Inspection was to be made for all defects and the defects tallied on a data sheet as in figure 2. If more than 3 defective castings were found the job was to be stopped and corrected. Samples of 10 were collected and inspected until a decision could be made according to multiple sampling procedure. The job could be accepted on the second sample if no defectives had been found.

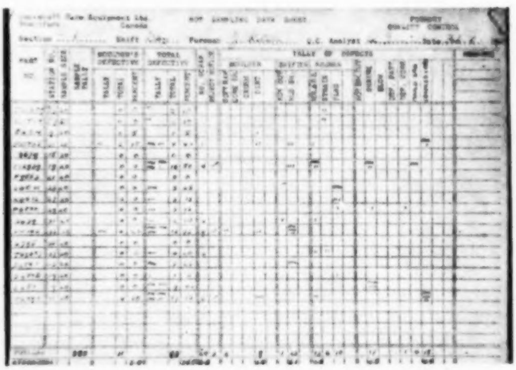


Figure 2

Then four more moulders were picked in the same section of the foundry, among whom were two good moulders. The procedure was repeated with one job from each moulder being sampled and inspected as above. Defects were tallied on the data sheet under the party responsible as far as was possible. A chart (figure3) was placed at each of the five moulding stations, and the percent defective which was the responsibility of each moulder was plotted on his chart. The difference between the good moulders and the poor moulders immediately became apparent.

When a job was rejected a quality control reject notice (figure 4) was issued to the foreman. This form was in triplicate, the third copy being retained by the quality control department. The first copy was retained in the foundry office and the second copy returned to the quality control department showing what corrective action had been taken.

At the end of the day the data sheet was totalled in all columns and

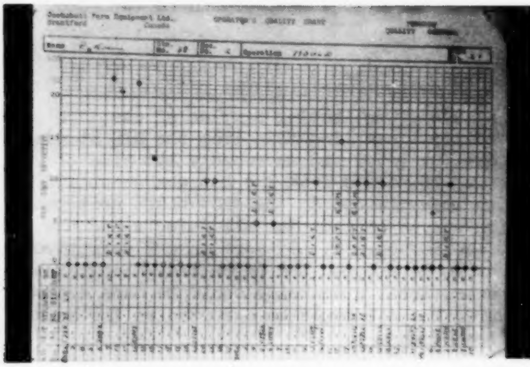
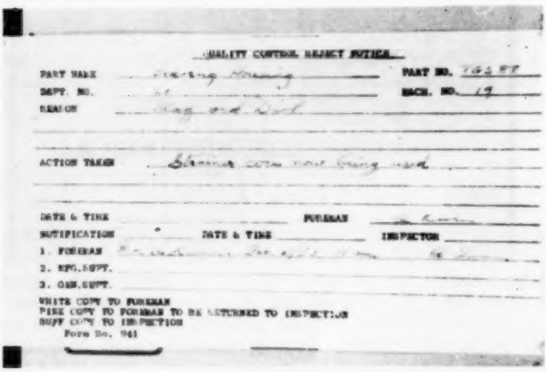


Figure 3



QUALITY CONTROL REJECT NOTICE

PART NAME ____________ PART NO. ______

DEPT. NO. ____________ MACH. NO. 19

REASON ____________

ACTION TAKEN ____________

DATE & TIME ____________ FOREMAN ____________

NOTIFICATION ______ DATE & TIME ______ INSPECTOR ______

1. FOREMAN ____________
2. MFG.SUPT. ____________
3. GEN.SUPT. ____________

WHITE COPY TO FOREMAN
PINK COPY TO FOREMAN TO BE RETURNED TO INSPECTION
BUFF COPY TO INSPECTION
Form No. 941

Figure 4

averages calculated for the five moulders, the five jobs including all defects, and the most prevalent defects. The data sheet was shown to the foremen and foundry superintendent the next morning to give them the information necessary for concentrating attention on the worst moulders, jobs, and defects.

The procedure was gradually spread over the entire two sections producing light castings. One major change was made for economy. It was found that 95 percent of all decisions were made after collecting two samples of 10 castings. Therefore, 20 pieces of each casting were collected at one time. The collector was instructed not to pick up any two

consecutive pieces of any small casting appearing over the hot conveyor, so that more than one mould would be sampled in the case of multiple moulds.

In the third section of the foundry where all heavy castings were produced which could not be handled by conveyor belt, hot inspection was introduced to cut down the time required to find and report the obvious defects producing scrap. Items such as shrinks, heavy scabs, misruns and strains could be seen easily at this time, and reported on a quality control reject notice. Because of lower production rates only 10 pieces of each job were set aside in a special cooling area for priority handling. These samples were cleaned in the shot blast unit as soon as they could be handled, and submitted for inspection. The results of this inspection were tallied on the data sheet as with the small castings.

A steady improvement in quality of moulding became evident early in the program. The good moulders were getting zero percent defective; the poorer moulders were approaching zero percent defective; the bad moulders stood out like sore thumbs and their foremen reinstructed them or changed their jobs.

What was more important, the moulders drew to the attention of the foundry supervision any defective cores, moulding sand, or pattern equipment. As a result the percent defective was reduced from 35 to 16 inside of a year's time.

Control of Core Quality

To assist in the prevention of scrap due to defective cores, an inspector was placed in the core room to view all outgoing loads of cores and to inspect as many cores in process as possible. The 4% double sampling plan in the MIL Standard 105a was used for inspection of cores ready for the foundry. Spot inspection was considered sufficient for most cores in process because very little trouble originated with the core makers. A poor sand mix or a defective core box would show up in a spot inspection of five cores.

For a few of the more costly cores percent defective charts were used and the plotting of them was based on the defectives found after baking or during assembly into the final core assembly. The greater amount of scrap proved to be due to careless handling, and the psychological effect on the handlers was such that this type of scrap was reduced rapidly.

Cupola Control

Defects attributed to the quality of the iron, and beyond the control of the moulder, pourer, and mould shifter, were countered by establishment of a cupola control based on chill depth, spout temperature, carbon content and silicon content.

The chill depth was measured in the broken cross-section of a small wedge of iron (figure 5) poured in a baked sand mould open at the bottom, which was placed on a square carbon block, using iron taken from the forehearth or mixing ladle under the cupola spout. This depth was measured in 64ths of an inch using a standard 6-inch scale, and together with readings of spout temperature, it provided the means of detecting quickly changes in the iron.

As stated before, the iron poured in our foundry is designed for 29,000 to 35,000 lbs. per sq. in. in a tensile specimen made from a cast bar 1.200 in. in diameter. The analysis desired is as follows:

Carbon	3.30 to 3.50%
Silicon	1.90 to 2.25%
Manganese	.50 to .75%
Sulphur	.15 max.
Phosphorus	.20 max.

This iron was inoculated with ferro-silicon or ferro-chrome as it entered the pouring ladles, according to the indications of the chill measurements and the size of the section and physical requirements of the castings to be poured. Thin sections where there was machining to be done received iron with higher silicon content. High strength castings received iron with practically no inoculation. Shares received iron with ferro-chrome additions.

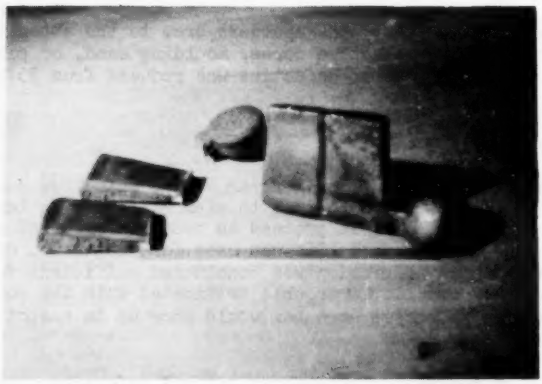


Figure 5

The chill depth measured at the cupola was controlled between 15 and 32 64ths to produce iron within the desired strength limits, and in a machinable range, without the necessity of large additions to the ladles which would make the iron cooler than desired for thin sections.

$\bar{X}$ and R charts were set up at the cupola for chill depth, spout temperature, carbon and silicon. Averages and ranges were plotted from three successive tests (figure 6). Tests for chill depth and spout temperature were made every 15 minutes. Carbon content was determined in the chemical laboratory every half hour. Each plotting of average and range was made from the last test value and the previous two. In this way the effect of normal variations in the cupola were not accentuated and the trend of operation was more obvious.

Corrective action was based first on the chill depth, the necessary amount of ladle inoculation being changed as indicated, and this was

followed by an immediate check up on the cupola charges. If the chill depth persisted in an out of control trend, an immediate check on carbon content was usually enough to verify any need for a change in the charging instructions due to some unexpected change in the cupola conditions or charging materials. The charts helped the cupola foreman to detect errors in charging by his men or a change in the coke or the scrap in his charges.

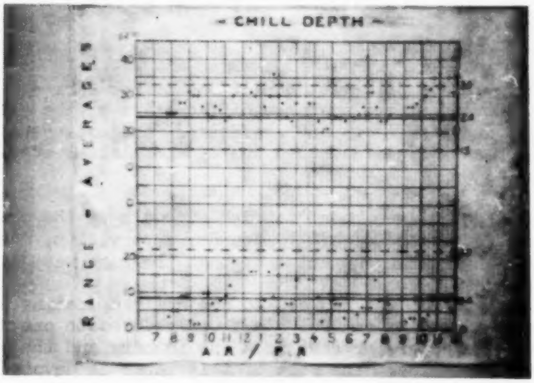


Figure 6

The charts were made to accommodate the 17-hour interval of two shift operation and were plotted on glass over the chart, using a marking crayon. If two cupolas were running, red and black crayons were used. Only at the beginning of each shift was the previous charting for that shift rubbed off. Thus a competitive spirit between shifts was encouraged.

Sand Control

Average and range charts of the same type as used for cupola control were set up for plotting of moisture tests. The tests were taken every half hour from each of the two sand mullers and four times a day from a hopper in each of the three moulding sections in the foundry. Permeability, green strength and deformation were plotted only in the sand laboratory, as these were controlled less by the operator than by return sand conditions and indicated to the metallurgist any need for change in his instructions for the mixing of the sand. However, the moisture charts were displayed for the muller operators and the moulders to see. The psychological effect on the muller operators was to bring the moisture into control; the effect on the moulders was to reduce dry sand complaints to the vanishing point.

Staff

With two eight-hour shifts operating, the quality control staff consisted of the core room inspector, a man at the cupola on each shift to

take readings for the cupola control and plot the charts, two labourers collecting castings on the dayshift and one on the nightshift, one checker or analyst, a supervisor over the collectors and analyst who also made any special investigations into causes of serious defects, and a quality control clerk to assemble data, compile reports, and follow up jobs reported defective but not corrected quickly. The efforts of this staff were co-ordinated by the foundry quality control supervisor in co-operation with the foundry superintendent and the chief inspector.

Present Procedures

New Conditions

With deterioration in market conditions in the farm equipment industry on this continent and exchange difficulties on the export market, there came cuts in production and layoffs and readjustment of labour throughout the plant. This affected the type of labour available in the foundry.

The rate of defectives began to rise and with it the weekly scrap. It was only with concerted effort by foundry supervision, inspection, and the methods engineering department, that the rise was halted.

When the production was reduced to one shift operating at about one-half capacity, the quality control staff was reduced to one test man at the cupola, one inspector for core room and foundry and the quality control supervisor part-time. The rest of this paper is devoted to an outline of the streamlined procedure developed to meet these conditions.

Hot Casting Control

Small castings are being checked by the moulding section foreman as they come over the hot conveyor belt, so that the more obvious defects may be detected with minimum delay. Samples of about five pieces of each casting are being cleaned in the shot blast unit and submitted to the foundry inspector. This procedure is repeated twice each day.

The simple check chart shown in figure 7 is placed at each moulding station. When scrap is found in the sample, the inspector marks the square opposite the part number and in the proper day column in red. When defectives are found which can be used, the square is marked in blue. When no defectives are found, the square is marked in black. The details of the defects found are entered on the back of the chart as shown in figure 8.

Large castings are being inspected for prominent defects as they are shaken out. Two or three per shift are being put through the shot blast unit as soon as they are cool enough to handle, and submitted to the foundry inspector.

The inspection results are summarized to give the following information: the percent defective and the percent coverage of the jobs poured (for the daily report to the chief inspector), and the total number of each defect found during the day, which is transferred to a weekly summary to highlight the most prominent defects.

The foundry supervision is the first to be notified of defects as the foreman is present during the inspection of his castings, and a

quality control reject notice is issued immediately when a major defect is found to occur in more than one piece in a sample. In the case of any scrap the job is stopped by the foreman until proper corrective action has been taken successfully.

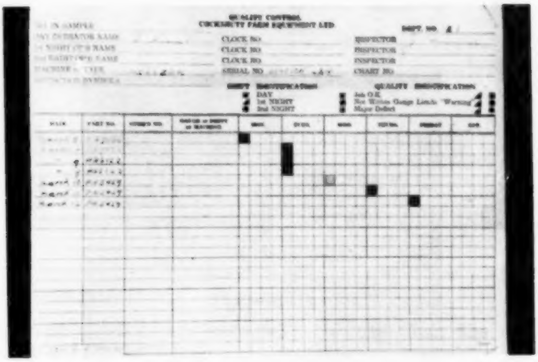


Figure 7

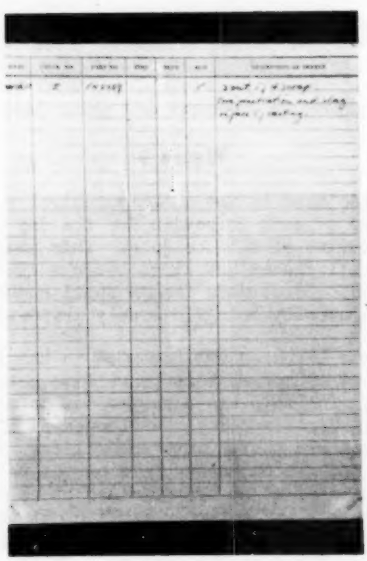


Figure 8

Control of Core Quality

The foundry inspector is required also to check cores in storage ready for issue to the foundry, using the 4% double sampling plan. The sample size for reduced inspection is used in all cases until a defective is found, whereupon the inspector reverts to normal sampling. All results are tabulated on the form shown in figure 9.

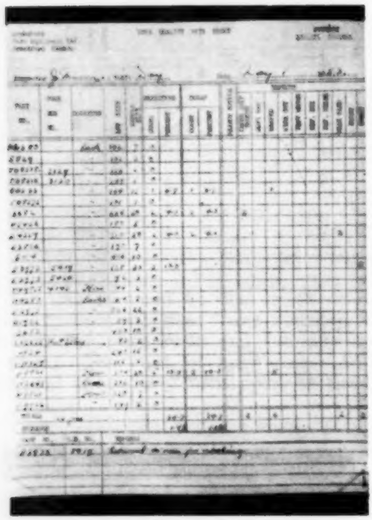


Figure 9

This form is also used by the quality control supervisor for summarizing for the average percent defective for the day based on the average of the percents defective of all the jobs. This average is used rather than the average of all pieces defective because it does not accentuate a large number of small cores more than a small number of large cores, the quantity of each usually being inversely proportional to the size.

All rejected lots of cores are reported to the foreman on the quality control reject notice so that preventive action can be taken when making more of them. The lots are scrapped out or sorted and repaired by the core packers. Certain cores are cleaned of fins as a standard operation. Some cores are pasted into assemblies. All such work is tallied by the core packers and assemblers on the tally sheet shown in figure 10. Thus counts of scrap are obtained at no extra cost. The inspector checks the sheets and enters remarks about the causes of scrap before showing them to the foreman and then delivering them to the quality control supervisor for a weekly summary. Unless the small cores are running 5% scrap or over, they are not included in the weekly summary nor reported to the cost department. This is considered a permissible scrap

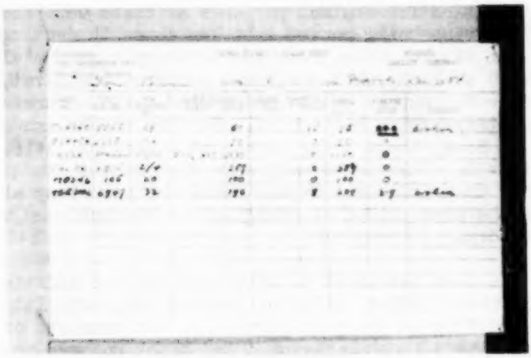


Figure 10

allowance as the cost of reporting would be more than the cost of the cores. However, all large cores are reported on a weekly basis to the cost department, the inspection department, and the foreman, as a standard inspection function.

Cupola Control

A new chart has been developed for cupola control which eliminates the need for arithmetic accuracy on the part of the chill tester. It is shown in figure 11, and consists of four histograms on one sheet $8\frac{1}{2}$ x 11. The distribution of the tests indicates definitely how the cupola is operating about one hour after tap out. As soon as the first three carbon checks are received from the laboratory, predictions can be made in regard to silicon content using the apparent operating means of the chill and carbon histograms.

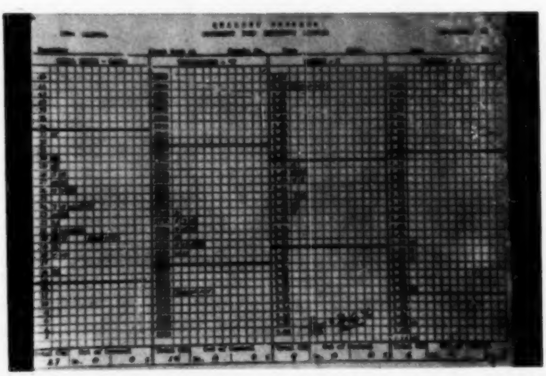


Figure 11

It was found to be impractical to take chemical tests in the first half hour of operation for control purposes as these were always high, the iron for that time being poured into low strength castings. Carbon tests are made on samples taken from the forehearth at 8 o'clock, 8:15, 8:30, 8:45, and 9 o'clock, with a silicon check on the first, third, and last. Thereafter the cupola can be controlled quite accurately using the chill and spout temperature histograms. Further checks are made after the shut down at noon, these being carbon at 1 o'clock, 1:15, and 1:30, and silicon at 1:30. This is to make sure that cupola conditions have not changed more than indicated by the chill depths. They also provide assurance that the same method of charging can be continued without change the next day.

A gadget has been developed to help regulate the additions of inoculant to the different ladles, without throwing responsibility on anyone but the pourers. This is shown in figure 12. It consists of a piece of plywood painted with a light background and black numbers in vertical sequence representing the chill depth in 64ths of an inch, hooks beside the numbers on which is hung a black metal indicator, and columns for the large and small pouring ladles and for plow share iron. Horizontal lines are drawn at various levels to enclose areas in the ladle columns showing the proper ladle addition for each range of chill depths.

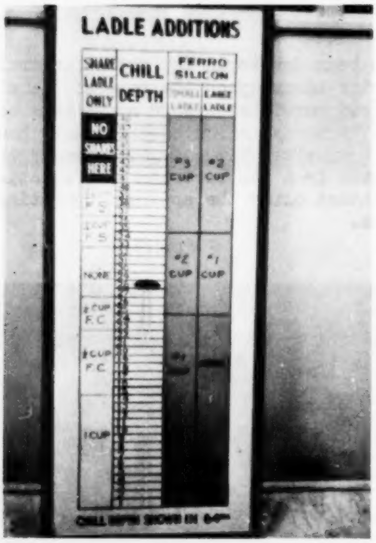


Figure 12

The position of the indicator is adjusted after each chill test to show the value of the test and automatically to indicate to each pourer the proper addition to be made to his ladle.

To ch[illegible] the inoculations, chill tests are made from the ladles at irregula[illegible] intervals, and four test bars are cast during the day, two from large ladles and two from small ladles. These bars are sent to the laboratory for a transverse breaking test and a Brinell hardness test as a quick check on the strength. Half the broken bar is machined to a standard tensile test piece .708 in. in diameter and pulled for a tensile strength value.

An index of the degree of control of the cupola each day is provided to the foundry superintendent by the simple process of dividing the number of plots outside the limits on the histogram charts by the total number of plots for the day. It is shown as a percentage, and is plotted on a chart in his office.

Sand Control

Samples of sand for test purposes are taken from the mullers and from the floor hoppers about four times a day. Tests are made for moisture, permeability, green strength and deformation.

The foundry supervisory staff has taken full responsibility for this control, and it will remain so as long as the castings made show good sand control.

Mill Room Control

The simple check chart shown in figure 7 is also used in the mill room to show whether the job being cleaned at each station is satisfactory or not. As each station is visited a sample of five castings is inspected and the results marked on the chart in the manner described earlier. Four to six inspections are made on each shift.

Defective work is drawn to the attention of the foreman immediately, verbally if possible. If the foreman is not immediately available, a quality control reject notice is made out and left on his desk.

Castings which are defective as received from the foundry and not previously reported are the subject of a quality control reject notice addressed to the foundry superintendent. Thus defectives missed by the sampling process at the hot belt conveyor are caught in the sampling in the mill room. There are very few reports of defective castings leaving the mill room and they are mostly for hidden defects revealed during a machining operation.

Reports

The chief inspector receives a report each day before eleven o'clock which provides information as to the percent defective in the foundry and mill room and the degree of inspection coverage.

The chief inspector and the foundry superintendent receive a weekly report giving full information on the average percent defective in the foundry, core room, mill room, and at the cupola, and a summary of details, causes and action taken.

Conclusion

In this paper, stress has been laid on the approach to quality

control. Full co-operation of the operators is essential to any quality control program. A slow start with full explanations to the union representative and individual operators, care in avoiding mistakes in assigning responsibility for defects, and readiness to correct any error (and there are bound to be some), made it possible to encourage operators to improve their workmanship from 10% defective to 0.5% defective measured over a month.

A slow start on rejecting jobs for being defective, with principle attention being focussed on jobs running high scrap, enabled the foundry management and the pattern shop to keep up with the necessary correction of equipment to eliminate the defectives, and thus their full support was maintained. It made it possible to bring along the development of quality control of the cores, system sand, and cupola iron concurrently with the work on the pattern equipment. This was important because lack of control in the sand or iron masked some defects caused by operator neglect or improper gating of the pattern. It was important also because the operators were encouraged by the corrective action being taken on sand and iron and cores.

Too much stress cannot be laid on the importance of the quality control reject notice. It is a demand for action which cannot be ignored by any foreman. It is a demand for action which can be abused by the quality control department by its unnecessary use, in which instance it can cause irritation and opposition. It must be used with good judgment.

To clear difficulties and maintain good understanding and co-ordination in all efforts to control of quality, a regular meeting between foundry management and the quality control department was found to be essential. It is held weekly or bi-weekly according to the apparent need and weekly reports and technical difficulties are discussed. Only through such meetings can full interest and effort be maintained. In our plant, one result was the persuasion of factory management of the value of a new type of flask equipment to reduce the number of defective castings with sand inclusions.

When production dropped off sharply and the quality control staff was reduced, some sampling procedures were stopped. It was soon found that the scrap rate began to rise. Analysis of scrap after the day's run was not sufficient to control it. Sampling, as described in this paper, was found to be essential for the maintenance of a low scrap rate.

Our management is convinced of the importance of quality control. One of our top officials said recently at a Foreman and Management meeting that he had been skeptical of quality control when it was started in our plant, but that now he knows that it is something our organization cannot afford to be without.

SOME USES OF STATISTICS IN PLANT MAINTENANCE

James Armstrong, Jr.
E. I. du Pont de Nemours & Co., Inc.

Introduction

It would probably be not too far from reality to say that many who are active in industry regard plant maintenance as a necessary evil, one which must be endured and which is being satisfactorily taken care of so long as production is not held up because of machine breakdowns. Evil or not, maintenance work is certainly costly as is evidenced by statistics from the chemical industry, where approximately ten percent of the employees are engaged in maintaining plants.

In the maintenance field, as throughout all branches of industrial effort, recent years have seen increasing use of the systematic study and evaluation of past performance for the purpose of finding means to accomplish future improvement. The use of statistical methods in such analyses seems to be coming more and more into play. It was a result of this general trend toward employing statistics to help solve industrial problems that led to the decision, at du Pont's Richmond, Va. rayon plant three years ago, to devote special effort to statistical study of our plant maintenance problems. I intend today to give brief descriptions of several specific instances where this work has proven valuable. I believe that these examples typify, to a certain extent, the fascinating and remarkable versatility of statistical tools.

Each of the examples which I shall discuss occurred in conjunction with the manufacture of viscose rayon yarn, a process which involves dissolving cellulose xanthate, a product formed by the interaction of cellulose, caustic and carbon disulphide, into a caustic solution and then coagulating the cellulose back to a solid in the form of multifilament yarn. In one process, subsequent to its generation, each rayon thread is built up into a cake inside a rotating spinning bucket. These spinning buckets are mounted on vertical spindles of small electric motors which turn them at several thousand revolutions per minute. Centrifugal force causes the yarn to adhere to the bucket wall and form a cylindrical cake as it is fed into the bucket. The relatively high speed of bucket rotation puts a desired twist into the thread as it is collected.

We have several thousand of these high speed motors in operation, replacing and repairing those which burn out or break up involves so much of our maintenance time and money that considerable effort is devoted to improving motor performance. How has statistics entered into this? Here are some examples:

I. The Chi Square Test

A. Two general types of motors are used. These motors differ essentially in the methods employed for dampening vibration. One type has internal provisions for dampening vibration and is rigidly mounted to the spinning machine frame. The second type is mounted on a flexible pad for vibration absorption.

In a certain period the motor positions in one of our spinning areas equipped with the two type motors had winding burnouts as follows:

Type Motor	No. Burnouts
Internal vibration compensation	68
External vibration compensation	406
	474

Since twenty-five percent of this area is equipped with the internally compensated motor, it is possible to calculate how the 474 burnouts would have divided between motor types, had each type performed equally, by apportioning that percentage to the internally compensated, the balance to the externally compensated type. Calculated performance is then compared with actual performance by means of Chi Square, the statistical coefficient of dispersion.

Frequencies of Burnout Occurrence

Type Motor	Portion of Area	Actual (Fa)	Calculated (Fc)	Chi Sq. $(Fa-Fc)^2/Fc$
Internally compensated	25%	68	118.5	21.3
Externally compensated	75%	406	355.5	7.1
	100%	474	474.0	28.4

Chi Square tables show that the probability of obtaining a value as large as 28.4 is less than .0001, which is to say in this case, the odds against chance occurrence of difference in performance, from motors of equal capability, as wide as that above are greater than 10,000 to 1.

This could mean that vibration absorption, or rather the lack of it, is responsible for the relatively high burnout rate of the externally compensated motors, a possibility somewhat bolstered by comparing the performance of those motors of this type recently mounted on spinning machine frames with that of the motors which have run continuously for some time:

Frequencies of Burnout Occurrence - Externally Compensated Motors

	Portion of Motors	Actual (Fa)	Calculated (Fc)	Chi Sq. $(Fa-Fc)^2/Fc$
Recently mounted	37%	306	231.4	23.9
Run continuously	43%	100	174.6	31.7
	100%	406	406.0	55.6

Here again calculated burnout frequency on this type motor is based on the hypothesis of equal performance from the two classes of motors. Comparing actual and calculated frequencies produces Chi Square even more significant than that obtained in comparing the two different types of motors. It is possible that in the recently mounted motors, vibration is constrained and load is thereby increased so as to produce excessive burnouts, whereas with the motors which have run continuously there has developed sufficient loosening of the motor mounts to make more free operation and thereby less load in the motor.

Our motor repair costs will be reduced in the order of $50,000 per year if externally compensated motors can be made to perform as well as the internally compensated type. As an outcome of this statistical investigation a more efficient vibration absorbing mount for the externally compensated motor has been developed.

B. In making rayon an important feature of the process involves controlling the size (denier) of each thread produced. To assist in accomplishing this control, the amount of cellulose solution which goes into each thread of yarn is metered by a gear pump which has critical parts accurately made to one ten thousandth of an inch tolerances. Performance criteria for these pumps together with the tolerance limits within which they must deliver for proper control of yarn denier, have been established. All the pumps are tested at regular intervals and those not performing within tolerance limits are replaced and overhauled at considerable expense. In several instances statistical study of pump performance and test data has indicated means of reducing pump overhaul costs. One of these instances serves as another example of Chi Square test's use:

In one spinning area two manufacturing operators are engaged full time testing those cellulose metering pumps. It was suspected at one time that one of these men was, through a slight departure from standard test procedure, imposing a bias upon his test results which could result in faulty condemnation of pumps with resulting inflation of replacement and overhaul costs. Records of an equal number of tests made by the two men in the same period of time showed:

Frequencies of Pumps Failing Test

	Actual (Fa)	Calculated (Fc)	Chi Sq. $(Fa-Fc)^2/Fc$
Operator A*	99	79.5	4.8
Operator B	60	79.5	4.8
	159	159.0	9.6

* Suspected to be using wrong procedure.

Here Chi Square indicates that the two operators could, from the same population of pumps, with the same test procedures, obtain results as different as those exhibited only about twice in 1000 series of tests. This data constitutes strong evidence that the difference in testing routine was affecting test results. When these facts were pointed out to the Plant Manufacturing Group, differences in test routine were eliminated and the results from the two operators became comparable.

C. A final example of Chi Square's use developed in studying spinning bucket motor failures where it was suspected that motor performance was adversely affected by the frequent breaking of funnel shaped guides which introduce each thread of yarn into its bucket. In this case motor failures in four equal periods were compared with frequencies which would exist had the motors failed in proportion to funnel guides breaking in each period:

Motor Failure Frequencies

Period	Actual (Fa)	Calculated (Fc)	Chi Sq. $(Fa-Fa)^2/Fc$	Funnel Guides Broken
1	117	115.5	0.02	190
2	93	92.5	0.00	152
3	89	92.5	0.13	152
4	97	95.5	0.02	157
	396	396.0	0.17	651

The statistical tables tell that chance alone will cause Chi Square

to be a larger value than 0.17 ninety-nine per cent of the time. Therefore, something other than chance must be causing motor failures and broken funnel guides to follow one another in the manner exhibited. This has led to the analysis and control of guide failures as a means of improving motor performance.

II. Frequency Distributions

A. Frequency distribution study is another statistical technique which has been valuable in this work. In one instance, it was decided to reduce the tolerance band within which the cellulose metering pumps, mentioned above, should deliver, in order to improve yarn denier control. Since such a move could involve an increase in pumps to be overhauled because of rejection on the regular test, we set out immediately to establish an estimate as to just how much pump repair costs would be affected by the proposed tolerance reduction.

To arrive at this estimate a random sample of the test deliveries of 344 pumps was obtained. The shape of the distribution of these measurements, indicated in Fig. 1A, was such that it was assumed to be normal. From the sample data the mean and standard deviation of pump deliveries were calculated and tolerance limits (existing and proposed) determined in units of standard deviations from the mean. Referring this information to a table of areas under the normal curve led to the conclusion that the proposed tolerance change would increase pumps to be overhauled by 15.8% of pumps tested. Fig. 1B attempts to illustrate the procedure followed. The translation of evidence into more generally understood terms means that the adoption of the proposed tolerance would increase pump overhaul costs by $15,000 per year.

Fig. 1A
Pump Delivery
Distribution

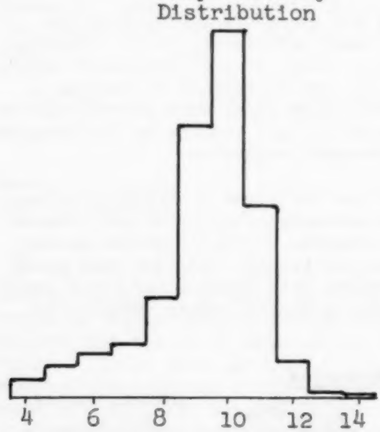
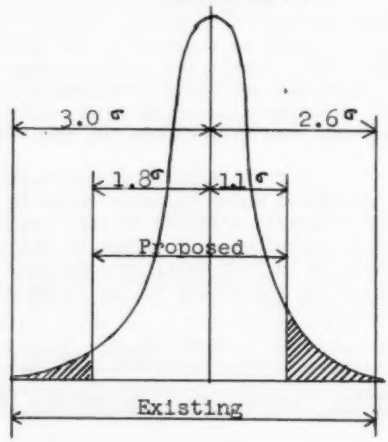


Fig. 1B
Pump Delivery
Tolerances

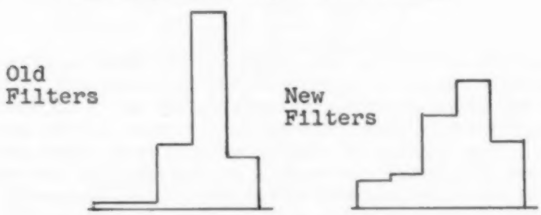


B. This case illustrates, I believe, a concept which has been referred to as the statistical point of view. Many of our technical people were of the opinion that the narrower tolerance band would involve an increase in the number of pumps failing to meet the specifications only until each pump in the plant had been tested once under the new tolerances.

They felt that, after the ensuing replacement and overhaul of those testing out of tolerance, only those pumps which became worn out would deliver outside tolerance limits on subsequent tests. This would mean that the shape of the test delivery distribution would then be altered to essentially that of the part of Chart 1B lying between the two shaded areas, an implication that the extent to which pumps are worn is the only factor affecting their performance. The statistically minded individual, on the other hand, suspects that pump condition is only one of many factors, all of which will affect pump delivery rate. There is considerable evidence, as a matter of fact, that such variables as viscosity of the cellulose solution being pumped, together with pump suction and discharge pressures have much more pronounced affects on pump performance than does pump condition.

In any event, presenting this information and point of view served to stave off any tolerance change but the question still remained: Why do the pumps deliver so erratically in this particular area when the same type pumps perform satisfactorily within the narrower tolerance band in another area? On investigating this it was found that different test procedures were followed by the two areas: where the pumps performed well, they were always tested when delivering through old filters; in the area where pump delivery was more erratic, tests were made with pumps delivering through new filters just installed. Apparently installation of new filters brings about a wider dispersion of pump deliveries, probably because resultant decrease in pump discharge pressure causes some pumps to deliver more, while incompletely filled filters, at the time of testing, result in other pumps delivering less. The data shown by Fig. 2, deliveries of a group of pumps with old and new filters, supports these suppositions:

Fig. 2
Effect of Filter Replacement
on Pump Delivery Distribution



Ratio of Std. Deviations = 1.5

As a result of these findings identical test procedures and tolerances are now used in the two plants with no increase in repairs.

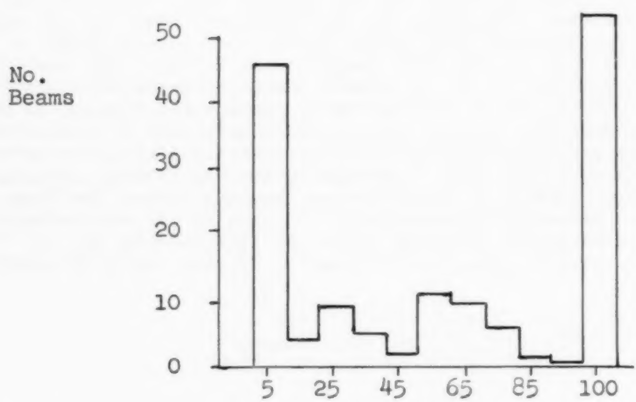
C. One of our textile areas in which yarn is finished and prepared for shipment to our customers supplies another example of gain through frequency distribution study. There the yarn is run through an operation called slashing, which involves sizing, drying and finally winding the yarn onto a large spool which contains several hundred pounds of yarn when filled. In this spooling operation, the spools, or beams as they are called, which rotate at fairly high speed, frequently commence to run erratically, sometimes to the extent that yarn being wound on them is

damaged. When this occurs, the operation is stopped and beams are doffed short of their intended load with resultant decrease in operating efficiency and increase in cost due to reprocessing involved. Poor condition of the beams and of machines on which they are run are felt to be responsible for these rough running beams. Therefore, considerable effort goes into attempts to keep poor running beams at a minimum.

In studying this problem, a group of 147 beams reported as having run erratically during slashing were classified in accordance with the percent of full load on them when doffed, this data is shown by Fig. 3.

Fig. 3

Rough Running Beams Classified
By % Loaded When Doffed



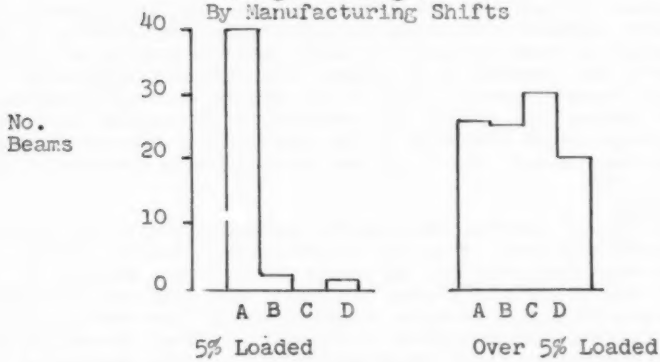
% Loaded at Doff

A distribution of such an odd shape as this will usually have some logical explanation, such is believed to be the case here. First it was believed that the large quantity of beams doffed at 5% of their load was caused by the relatively rough running of all beams at the start of the slashing cycle. The barrels of these spools in some cases are not smooth, as a consequence the spools run roughly at the start of the operation and it requires considerable judgment to estimate which beams will settle down with yarn build-up and which will continue to run poorly.

The equally large quantity which was allowed to fill completely, even though running poorly, is believed to be made up of beams which never behave quite bad enough to warrant pulling off the machine before completion of the run. The remaining portion of the distribution occurs in a rather normal fashion, smooth tapering to the extremes with the maximum quantity occurring essentially in the center.

With this information, it was decided to compare the four manufacturing shift crews with one another, particularly with regards to those beams doffed at 5% of intended load since there, as has been mentioned, the exercising of considerable judgment is required. Fig. 4 shows this classification of the data.

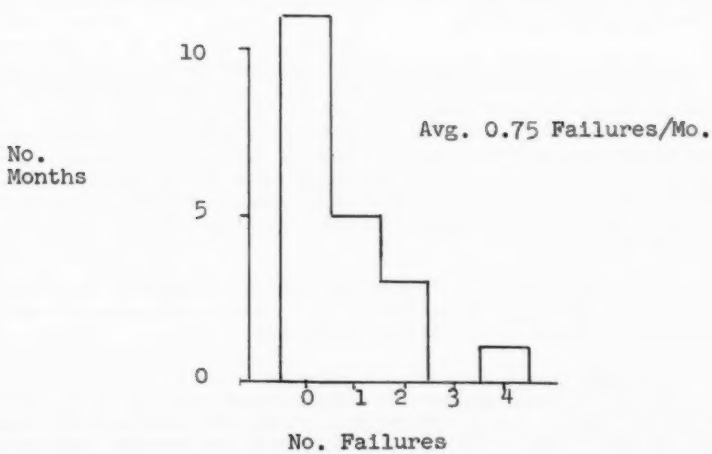
Fig. 4
Rough Running Beams -
By Manufacturing Shifts



Chi Square was not needed to tell here that "A" shift was significantly high with respect to beams doffed at 5% of load while no better than other shifts in other respects. As a result of this study, there has been an almost complete elimination of beams stopped and doffed early in their cycle because of their poor running performance.

D. For a final example involving the use of frequency distributions, some time ago there developed considerable demand, due to excessive failures, for redesign and replacement of bearings in the electric motors driving machines on which one of our textile operations is performed. Failures of these bearings during the preceding twenty months were distributed as indicated by Fig. 5.

Distribution of Bearing Failures
in 20 Consecutive Months
Fig. 5



Practically all the demand for redesign stemmed from the high failure period.

Reference to tables of the Poisson distribution, applicable to data of this sort, revealed that there was a theoretical probability of .0074 of obtaining 4 or more failures in a month when the average per month is 0.75. While this probability is slight, it is not so slight as to warrant its being considered proof of the need for a $10,000 bearing replacement program. In view of this evidence, we recommended that the bearing change not be considered on the basis of the evidence at hand. The change was not made, there has been no occasion to consider it since.

III. Sampling

As in Quality Control, sampling for percent defective has proven useful in maintenance work. This will be indicated by one of a number of instances where this technique has proven useful. In a program to recondition some of our purification equipment, when cost and continuity of production prohibited accurate determination of just what quantities of many different equipment parts were needed to effect the reconditioning, parts requirements were estimated through detailed examination of a random sample of equipment pieces. It seems that maintenance men are prone to overestimate material requirements on large scale overhaul jobs. In this particular program, on evidence furnished by the statistical sample, parts costing $31,000 were ordered. Upon their receipt some six months later they were adequate, were all used and served to put the equipment in satisfactory condition.

IV. Data Requiring No Analysis

Another benefit worth mentioning in this paper is the savings sometimes derived through just the orderly collection of data. The following are two examples of this sort of thing. The first concerns one of our regularly scheduled maintenance overhaul jobs which involves removing an item of equipment from a manufacturing machine and replacing it with a like item which has been cleaned. A simple investigation revealed that this job was costing $60,000 per year, of which a major portion was the cost of equipment unavoidably broken in making the change. This cost figure, together with knowledge of the process, was sufficient to persuade all concerned that the overhaul frequency could be halved without affecting product quality.

A second example of gain through the mere collection of data occurred in an area where low machine efficiency was having an adverse influence on productivity. Data collected to determine lines along which desired improvement in mechanical efficiency might be obtained revealed that breakdowns were causing a 5% loss of machine efficiency. This, coupled with the fact that overall efficiency of the machines involved was only 60%, clearly indicated that no amount of money spent on improving maintenance would appreciably help the manufacturing group in performing its task.

While admittedly not derived through the science of statistics, I think that benefits such as these will accrue in any program where facts expressed in terms of figures are collected.

V. Correlation

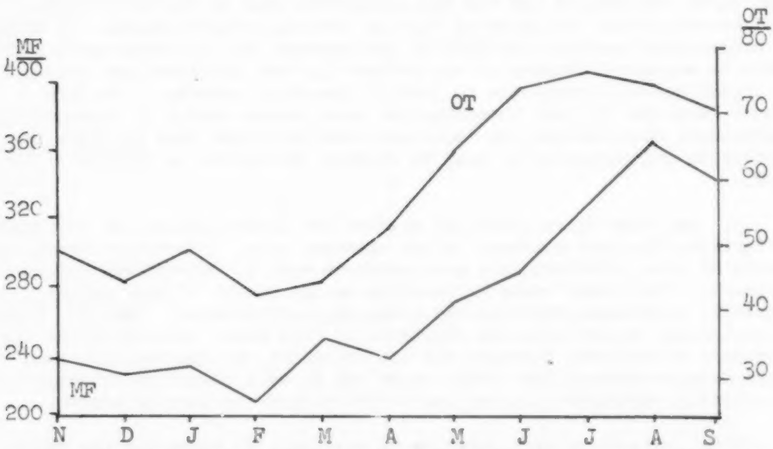
A. Correlation analysis is a statistical tool which has proven extremely useful. After considerable study, the relationship between spinning bucket motor failures, in one area, and outside temperature,

shown on Fig. 6, was discovered and verified statistically.

Fig. 6

MOTOR FAILURES AND OUTSIDE TEMPERATURE

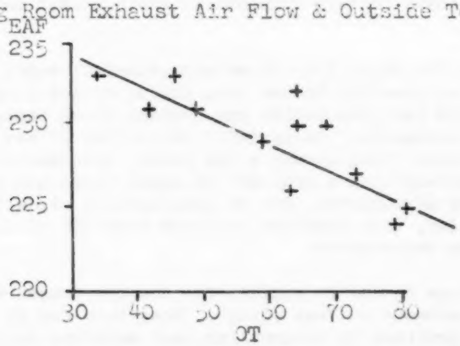
n = 11 r = 0.908



No apparent reason for this relationship was at first evident since the motors operate in an air conditioned room where temperature is fairly constant and there is but slight response to changes in outside temperature . Then it was postulated that a rise in outside temperature would result in a decrease in draft action generated in the plant fume stack. This decrease in exhaust air flow might result in a decrease in cooling air exhausted across the motors into the fume exhaust system. This supposition was verified by taking exhaust system air flow measurements at varying outside temperatures and establishing the fact that a relationship existed between air flow and outside temperature as shown on Fig. 7.

Fig. 7

Spinning Room Exhaust Air Flow & Outside Temperature



n = 15 r = -0.818

Measurements of motor temperatures showed them to be highest where air flows were least. Finally, when temporary circulating fans blowing on bucket motors brought about a decrease in failures of motors so treated, installation of permanent fans to circulate air over all the motors effected a considerable reduction in motor failures.

B. Several thousand units of one of our important production equipment parts had been in use for two years when some of the units began to fail because of an undiscovered flaw in internal reinforcement. By means of correlation analysis the loss in the average life of these parts which would be sustained because of the design flaw was predicted and proved to be of sufficient consequence to justify immediate redesign. Failure trends developed in this investigation have proven useful in forecasting replacement requirements, an important item since the part is vital to production and replacements must be ordered six months in advance of their need.

C. Multiple correlation was used in the investigation for the cause of variation in yarn moisture in our slashing area. These "wet beams" or spools of yarn containing too much moisture must be reprocessed before shipment. "Wet beams" were believed to be the result of poor drying caused by condensate flooding the steam drying cylinders. However, when investigation showed that the frequency of "wet beam" occurrence and the frequency of cylinder flooding did not correlate, it was realized that other factors entered into their cause and we were therefore reluctant to institute expensive alterations to the slasher condensate system.

Final control of yarn moisture is exercised by comparing the weight of the yarn on each beam with its theoretical weight at desired moisture content. Beams are shipped if yarn weights are within certain limits each side of theoretical, out of limit beams are reprocessed.

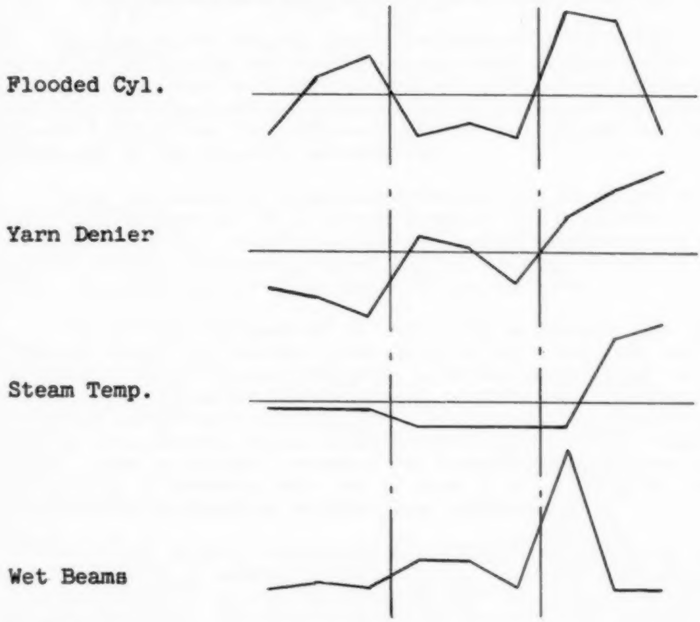
Multiple correlation showed that yarn denier variation played a large part along with flooded drying cylinders in causing "wet beams" to occur. A third factor, temperature of the drying cylinder steam, was found to have an effect about equal to the effects of denier and cylinder flooding in causing "wet beams".

Fig. 8, presenting a highly condensed version of the data which was analyzed to obtain these conclusions, is intended to give a general idea as to how these three factors combined in their effect on the occurrence of "wet beams".

By dividing the chart into three approximately equal chronological periods, it may be possible to see that during period 1 cylinder flooding was high but, with low yarn denier and nominal steam temperature, "wet beams" were not excessive. In period 2 the number of "wet beams" increased even though cylinder flooding was a low level, presumably this resulted from a denier increase and a drop off in steam temperature. In period 3 "wet beams" first skyrocketed, due to simultaneous rises in cylinder flooding and denier, then subsided with the drop off of flooded cylinders and rise in steam temperature.

These findings tended to explain the obscure causes of wide variation in the occurrence of "wet beams". They have led to a better understanding of our problems in controlling yarn moisture and to the decision that expensive redesign of drying cylinder condensate removal equipment is not required to obtain this control.

Fig. 8
Wet Beams and Related Process Variables



<u>Conclusion</u>

In outlining these various instances in which statistics has proven useful in studying maintenance problems, it is hoped that you have gained some conception as to how this science can be applied to any field where facts can be expressed numerically.

THE TEST PILOT'S ROLE IN QUALITY CONTROL

John I. Nestel
Aviation Engineering Division of Avien-Knickerbocker, Inc.

The segment of industry which participates in the aviation field is well indoctrinated in the extensive qualification requirements and standards that must be satisfied. The aviation industry must not only meet the rigid specifications designated by the Bureau of Aeronautics, the Air Force or the C.A.A., but in addition must comply with the requirements of the aircraft manufacturer.

These voluminous detailed specifications are the ground rules which today are regarded as fairly common practices to be followed by the aviation industry. Therefore, it is well established that the extensive quality control programs prevalent in even the smallest aviation industrial organizations are fully justified and necessary.

The critical applications of most aviation products have resulted in a higher direct and indirect labor ratio of inspection and quality control personnel. Manufacturing facilities producing aeronautical equipment that affects the safety and performance of an airplane have no doubt accelerated management's unquestionable recognition of the advantages derived from placing Quality Control Departments at a top organizational level. This, of course, precludes the possibility of adverse effects on the quality of products which may be encumbered by the normal objectives of the Sales, Engineering or Production Department.

The final quality evaluation of the entire aircraft, its components and accessories is made by the Test Pilot. Up to the test flight stage all the qualification test requirements have only been performed on components, equipment and structures during the prototype stage or on representative samples of production units. During initial flight test, the Test Pilot brings the entire aircraft into the actual physical and environmental conditions, many of which have previously been simulated individually during the qualification testing program. The extreme ranges of accelerated gravitational forces—temperature, vibration and pressure—are reached during test flight, and the Test Pilot then becomes, in the eyes of Quality Control, the "Airborne Inspector." He checks the entire aircraft and its systems to determine whether the performance, reliability and safety standards have been satisfied.

The Test Pilot's professional qualifications in both flight testing and technical knowledge enable him to adequately evaluate the various quality phases of an airplane and its components. He then follows on a different level the job of the inspector attached to some quality control section in the manufacturing process, who uses test equipment, designated standards or simulated conditions to evaluate the operation of individual components or installed systems on an aircraft. The role of the Test Pilot becomes the last step in the methods or operation sheets of the Quality Control Department, with the responsibility of evaluating quality under actual flight conditions.

The Test Pilot should be able to render the same non-partisan decision that is made by the inspector who is part of a quality control organization with his channel of authority answering directly to top management. Most large aircraft manufacturers have recognized that the Test Pilot's role in quality control must be enhanced by providing the

proper system of quality control coordination and liaison facilities in obtaining maximum benefits from the flight test organization.

The results of the Test Pilot's evaluations and decisions are awaited with great anxiety by those who were responsible for designing, building and coordinating a quality product for all specified flight conditions.

We are very fortunate to have on our panel three well experienced Test Pilots. They are Mr. Richard Taylor, Project Test Pilot, Boeing Airplane Company, Wichita, Kansas; Mr. R. C. Little, Chief Test Pilot, McDonnell Aircraft Corp., St. Louis, Missouri; Mr. J. Angelone, Staff Experimental Pilot, Chance Vought Aircraft Division, Dallas, Texas.

IMPROVING VENDOR QUALITY PERFORMANCE

David A. Hill
Hughes Aircraft Company

Major airframe programs can involve sub-contracting fifty per cent of the plane's structure to the tune of 200 million dollars or more. This is in addition to the long inventory of electric motors, actuators, controls, servos, black boxes - either purchased or Government furnished - that the prime contractor must fit into his airframe; and the raw material, hardware and detail parts that feed his assembly lines. It is easy to see that the performance of his vendors becomes vital to the manufacturer's program; his profit or his loss can depend on how skillfully he handles procurement.

Not all procurement programs are on this scale, but another trend has been affecting manufacturers of advanced equipment. As Howard T. Lewis states in the Harvard Business Review,[1] "The more sales depend on excellence of design and high technical performance, the greater is the task of maintaining an effective balance among the procurement, the engineering, and production functions. Many managements have learned the truth of this the hard way. Trying to resolve the apparent conflict of interests that arises, for example, when engineers insist on the purchase of a specific item regardless of cost and delivery considerations is but a surface indication of a much deeper problem".

Difficult technical requirements force the manufacturer to beat the bushes for vendors who can accomplish the impossible. John A. Cairns in Printers" Ink,[2] points out this reversal of roles:

"Traditionally, business has looked to its customers as its source of profit. The reasoning was simple. You sold at a profit, and the profit came from your customer. That's all there was to it. You considered that your supplier had the same view toward you; that you, as his customer, were his source of profit, and that just as you courted the favor of your customers, he would court your favor. You addressed all your merchandising, promotion and advertising efforts at your customers, and your supplier addressed his efforts at you. Every selling avenue had a one-way sign."

"Today, however, science has altered the situation. Almost every selling avenue is now a two-way thoroughfare. While the customer is still primary, a new element has appeared in the situation. Today, the source of supply is in many cases a primary source of profit opportunity, just as the customer is a source of direct profit. Today, the business that is dependent upon scientific and technical ingredients for its products must direct a major part of its attention to the sources of these ingredients. It must maintain a sound position with its sources. Competition is keen for a favorable buying position with many basic suppliers."

When this session, sponsored by the Aircraft Technical Committee, was planned, Mr. J. Y. McClure and I felt that it would be interesting to poll the manufacturers in the field. A rather lengthy questionnaire was sent to 72 Quality Managers and Chief Inspectors. It asked about their current experiences, practices and problems in receiving inspection and vendor quality performance. Fifty answers were received --- convincing evidence of interest in the subject.

The fifty answers include a large majority of the airframe and engine companies. The auxiliary equipment and electronic manufacturers are well represented. In compiling a report on the results, the geographically separate divisions of large companies were counted independently, since it was obvious that they had arrived at individual answers to the questions raised.

The questions were not only answered; many provided such interesting comments that most of this paper is based on the information received. In order to avoid any possible embarrassments, names of companies or individuals have been omitted, even though this does an injustice to the many who put real effort into providing valuable comments.

The quality men made it plain that their biggest problem was, "How do we improve vendor quality performance?" The cost of source and receiving inspection is in inverse ratio to this performance. The quality of their end products is strongly affected by it.

We asked their opinion on what caused rejections. The results are tabulated below:

		Ranking Frequency 1=most important cause, 2 next, etc.						
		1st	2nd	3rd	4th	5th	6th	7th
(a)	Inadequate or confusing design information	2	1	5	16	10	2	
(b)	Delays in getting design changes to vendors		4	7	7	11	7	2
(c)	Unauthorized personnel			1	1	4	8	7
(d)	Inadequate processes or equipment at the vendor	4	8	16	6	3	3	1
(e)	Errors by vendor's operators	11	20	7	1	5	1	
(f)	The policy of accepting the lowest bid	1	1	8	7	2	4	10
(g)	Lack of consistent inspection or quality control by the vendor	29	13	3	2			

The lefthand side of Figure I shows these results in histogram form.

The vendor's quality control system takes a beating in the answers to the question (g). Boners by his operators (e), and poor equipment (d) are close behind. This may point up the urgency for more vendor education, which will be mentioned later. While (a) and (b) are not in

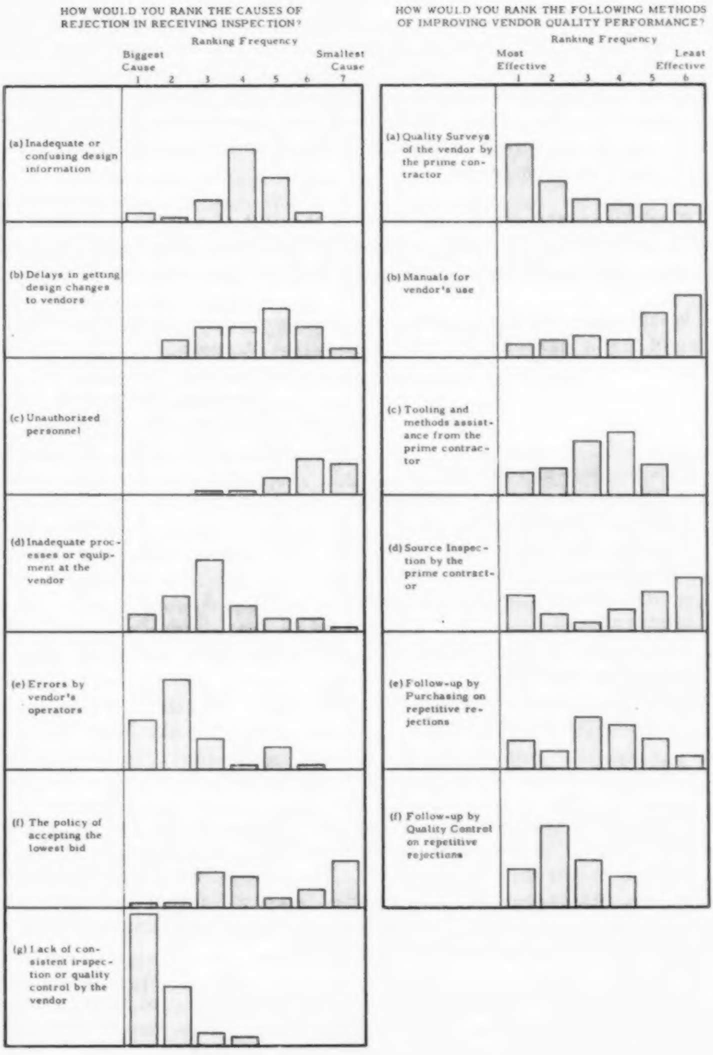


FIGURE I

the top ranking, many companies recognized that they sometimes confused or frustrated their vendors' efforts to please them. A clue to getting rid of this problem came from one company, "Items (a), (b),

(c) and (f) are not found to be contributing factors in rejections occurring in Receiving Inspection. This can be attributed to coordinated effort between Quality Control, Engineering, and the Materials Department. Design changes are coordinated with Change Control and the subcontractors before effectivities are listed. The Materials Department, as a matter of policy, does not issue a Purchase Order unless this Purchase Order is screened by Quality Control and an actual Quality Survey of the Vendor is considered".

Some words got left out of (c) in the list of causes. It was meant to cover instructions to vendors by unauthorized personnel. If this had been clear, (c) might have been rated higher as a cause of trouble.

Well, what can the quality man do to get at the causes of rejection? We asked for a preference rating on a list of methods. Here are the results:

		Ranking Frequency 1=most effective aid, 2=next, etc.					
		1st	2nd	3rd	4th	5th	6th
(a)	Quality Surveys of the vendor by the prime contractor	17	9	5	4	4	4
(b)	Manuals for vendor's use	3	4	5	4	10	14
(c)	Tooling and methods assistance from the prime contractor	5	6	12	14	7	
(d)	Source inspection by the prime contractor	8	4	2	5	9	12
(e)	Follow-up by Purchasing on repetitive rejections	$6\frac{1}{2}$	4	$11\frac{1}{2}$	10	7	3
(f)	Follow-up by Quality Control on repetitive rejections	$8\frac{1}{2}$	18	$10\frac{1}{2}$	7		

The right side of Figure I plots these answers.

Surveys of the vendor and reject follow-up by Quality Control come out ahead. Apparently we believe that, if a thing is to be done well, we had better do it ourselves. The places where one-half values show up resulted from advocating that Quality Control and Purchasing should get together on an approach to vendors. Purchasing follow-up and help with tooling are considered useful.

You will notice a divided opinion about the value of source inspection. A separate question on the subject was, "How much value do you place on company source inspection?" The answers:

(a)	Valuable for most purchased parts -- reduces or eliminates rejections in receiving inspection ...	9
(b)	Useful only for major sub-contracts ...	28
(c)	Too expensive for results gained, to be avoided if possible ...	16

While the companies who use it widely are in the minority, the value they place on it is given by one comment, "We have been in the process of expanding our field quality control force for the purpose of source acceptance of those items which lend themselves to such a program. This philosophy has enabled us to obtain greater flexibility in our receiving inspection, at the same time permitting a minimum fixed within plant receiving inspection force. It is our intention to further expand this concept because we feel it has numerous advantages . . . both for us directly and to the interest of the individual vendors".

Among those who restrict its use, one company lists the factors considered: "Source inspection is performed in the following cases:

1. High value product
2. Where process inspection is necessary
3. Special equipment required for inspection
4. Bulky items, difficult to inspect upon receiving
5. Large volume of material "

Two companies have come to interesting conclusions about the approach to source inspection, "A surveillance inspection of the type indicated in MIL-5923A is probably the most that can be performed and still control the product without duplicating, augmenting, or replacing a vendor's inspection. The type of vendor who requires more than this will be expensive to the prime contractor and the product questionable at best. The source inspector is not in the position to have complete information, or sufficient time, or to be supplied with necessary engineering and tools to perform the required functional tests to assure a sound product. The vendor's tests would usually be followed without the assistance available at our plant. After all, the best and cheapest inspection is in the integrity and knowledge of the man actually doing the manufacture, processing, assembly, etc.", and, "We do not use it except in rare cases on simple purchased parts. We send quality control engineers to our major suppliers to review their process operations and controls and send inspectors to the vendors' plants to inspect major items at source."

There are other ways to reach out beyond the factory receiving dock in an effort to increase the proportion of acceptable material. As one company puts it, "A very important factor which reflects upon the quality control of a vendor, is that of knowing the quality requirements of the prime contractor with whom he is dealing. Our experience indicates that both improved quality and better consumer-producer relationships are evident when there is available for vendor consumption, a specific detailing of the prime contractor's quality level requirements." We asked several questions in this field.

Have you negotiated purchase order agreements with vendors that include specified sampling plans to be used on shipments received?

(a) Done as a general practice	...	7
(b) Done in a few special cases	...	11
(c) Not done at all	...	19

Do your vendors supply you with data from the inspections and tests they perform prior to shipment?

(a)	Done as a general practice	...	5
(b)	Done in a few special cases	...	40
(c)	Not done at all	...	5

At present most of the data received covers chemical and physical tests, X-ray, Magnaflux and similar materials requirements. Some have arranged to receive functional test data to assist in tracking down the cause of trouble or to calibrate vendor test methods with those of the prime contractor. A desire to increase this practice was evident. Two comments are, "We heartily encourage data with the shipment from the vendor. This gives an immediate opportunity to check the measuring methods of the vendor vs. (this company). We have had only limited success in receiving this advance data", and, "We have received from a few of our suppliers copies of lot plots, histograms, and process data, which is very helpful in performing inspections such as we do at this facility".

Have you developed mutually accepted inspection check lists with your vendors?

(a)	Done as a general practice	...	5
(b)	Done in a few special cases	...	25
(c)	Not done at all	...	15

This practice is not as wide spread as check lists for internal use, but it is on the increase. The vendor may become aware of check lists through information returned with defects. Vendor-Contractor check lists came up in the sub-contracting of major air frame assemblies. For example, "(This company) does not coordinate each and every check list at the time it is originated. We do, however, provide applicable vendors with copies of these sheets on all parts with which they may supply us. Should there be any question as to the adequacy of the characteristics listed thereon, the check list will be considered for revision. Further, we hold periodic conferences with our suppliers, for the purpose of coordinating our inspection technique, acceptance criteria, and general policy".

Some work has also been done on vendor proprietary items, "Supplier representative and contractor liaison is very desirable when establishing check lists for vendor proprietary items. Check lists are generally established by Quality Control in conjunction with Engineering".

We have obviously just begun to develop the possibilities of these approaches. Some suppliers are also beginning to think in terms of telling their customers more about the product shipped. Tele-Tech Magazine [3] reports one case, "Although quality control has now been in service in many electronic manufacturing plants for ten years or more, it is surprising how few contracts between customers and vendors specify what quality is expected. According to Malcolm Young, Director of

Quality Control Dept., Erie Resistor Corp., only a few customers are fully aware of the quality control and sampling plans used in controlling quality for them. .... When Erie certifies a certain quality level to a customer, they tell by what means Final Quality Control Inspection accepted the material for shipment. They specify what type Sampling Plan was used (MIL STD 105A, Dodge & Romig, Army Ordnance, Columbia U, etc.) and also state the quality level of the lots shipped".

It is encouraging to see one of the giants of the automobile industry using a well developed system. As reported by Roscoe Smith, of the Ford Motor Company, in Tooling and Production,[4] "Now Vendor Quality Level Certification moves into prominence as a sort of by-product of this new and developing science of Quality Control. It fills a long-felt need which has pressed equally upon the suppliers and consumers of production parts for much too long a time. This is the simple need for more adequate understanding of one another in the entire matter of Quality. And the greatest single impediment to the development of this understanding has been the lack of facts. It has been the lack of a concrete, mutually established and fully understood Quality Goal. It has been the consequent lack of a consistent system for Quality evaluation, with respect both to the consumer's incoming product inspection and the supplier's outgoing product inspection -- traditional efforts at 100% Inspection notwithstanding".

Mr. Smith outlines seven steps leading to certification:

(1) Select the characteristics contributing significantly to quality.
(2) Classify the characteristics as critical, major, minor and separate.
(3) Negotiate acceptable quality levels for each class of characteristics.
(4) Agree upon the general control system the supplier is using or will use.
(5) Agree on periodic written evidence from the supplier.
(6) Conduct a pre-certification trial.
(7) Sign the certification agreement after the trial period.

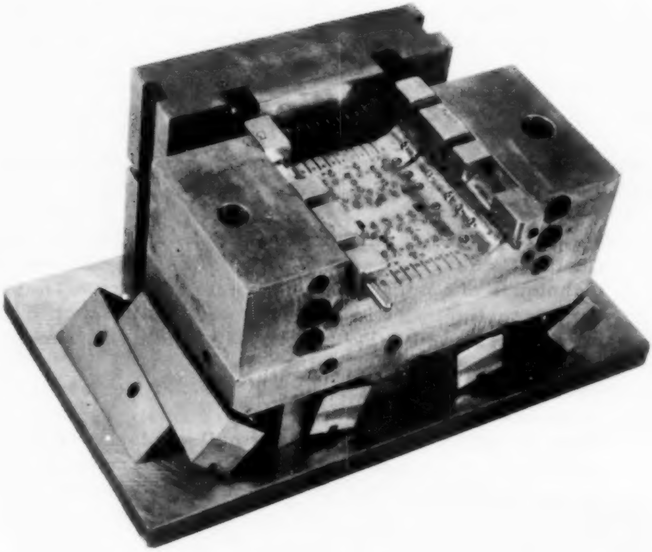
He further points out, "It is provided in our Certification Procedure that the Agreement, which becomes an attachment to the Purchase Order, should be signed by those the supplier might designate as responsible authorities, together with representatives both of Purchasing and Quality Control in our consuming division. The consumer thereby engages to accept and scrap or repair the defectives in any shipment which meets the test of the Acceptable Quality Levels. That is provided, of course, that any uncertified characteristics are not defective to a prohibitive extent. As soon as full dependence may be placed on the Certification, our own inspection efforts are reduced to a minimum. We continue on either a cycle or a random basis to perform "audit" checks insuring that the shipments furnished under Certification meet the claimed levels of Quality".

The quality men who answered the questionnaire frequently emphasized the need for personal contacts with the vendor. A typical statement was, "Our experience indicates the most effective method for getting corrective action is to work with the vendor at his plant, avoiding every possibility of misunderstanding, misinterpretation, etc., which can be created by remote communication". Most companies agreed that Purchasing has the primary responsibility for such contacts. But the quality engineer can work with Purchasing, using a specialist's knowledge to untangle problems. If necessary, the tooling and methods experts should be brought into the picture. Figure II shows our early experience with a group of difficult molded parts. The part produced by the mold shown in the picture turned out to be an excellent barometer for the entire group. The assembled parts, together with numerous terminals, connections, clips and circuit components, form a vital subassembly in a flying electronic brain. Before we got rejections down to where we could live with them, we had to help the vendor with his tooling, his process controls, his instrumentation and his inspection methods. The chart is only a reflection of the intensive effort that was necessary.

A glance at the first month's results might suggest that we would have been smart to get in the door a little sooner. This brings up the idea of Quality Surveys of the vendor. The industry rated surveys very highly as a means of preventing trouble before it gets started. For example, from an airframe manufacturer, "Most important of all is a thorough review of job by contractor from the point of view of planning, tooling and methods before the subcontract is placed". An electronics company agrees, "Field quality representatives who make the initial quality survey of the vendor, often follow up to improve quality performance". And one company comments on the previous listing of methods for improving vendor quality, "Items (a) through (f) inclusive are all valuable instruments in obtaining corrective action, however, it is felt that the requirement for corrective action would necessarily be held to a minimum if the following steps are taken prior to issuing a Contract:

(a) Quality Surveys and approval of potential sources.
(b) Quality requirements stipulated on the Purchase Orders.
(c) Frequent Vendor and Contractor Representatives conferences to discuss Tooling and Engineering problems, methods of test, and inspections.
(d) An adequate system of vendor notification of rejections within the shortest possible time."

Personal contact with vendors is necessary if the prime contractor expects them to understand his use of statistical methods in receiving inspection. Suppose your vendor has come to count on the fact that you screen his shipments and return only defectives. One fine day the whole shipment lands back in his lap. It did not meet your sampling plan, based on a newly established AQL. He will soon be burning up the wires to your purchasing agent. If the vendor is one of those hard-to-find producers of technically difficult components, the purchasing agent



BASE AND END PLATE PORTION OF "CLAMSHELL" MOLD

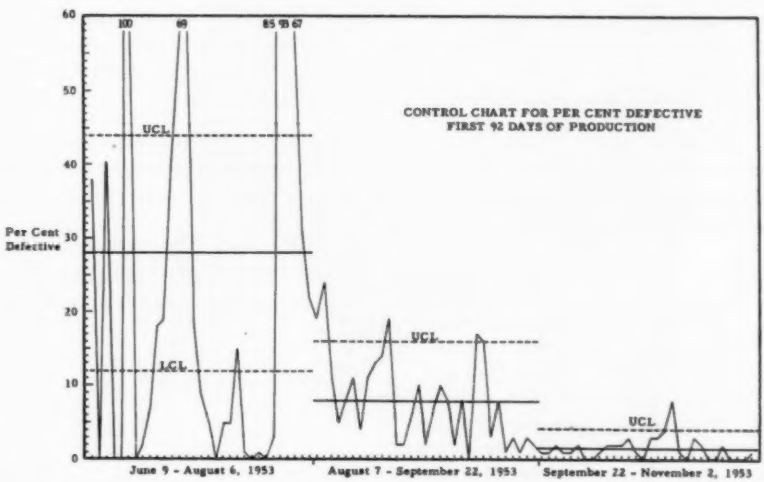


FIGURE II

is apt to be down on quality control for some time to come.

Our questionnaire asked how many had promoted the application of statistical quality control techniques with vendors. The answers:

Yes	26	Successfully	18
No	22	Unsuccessfully	2

About half the companies have tried this promotion. The results range from decided enthusiasm to modest hopefulness:

1. "We have conducted instruction courses and use our Vendor Quality Control Representative to continually promote the techniques of quality control with our vendors. They are very receptive. We have not, however, approached all of our vendors and believe that this is a field that should be explored more extensively."
2. "Promoted to a limited degree. Too early for evaluation of success."
3. "There is still a large group of small vendors who do not use S.Q.C. techniques and do not understand them. The larger vendors are rapidly becoming qualified in this field."

The hope was expressed that wider use of statistical quality control by vendors would result in reduced receiving inspection costs. One or two companies feel that the vendor can be allowed no leeway from the print requirements. Any discrepancies passed by his sampling techniques are subject to rejection and replacement. Others would dispute whether this policy is consistent with the Acceptable Quality Level concept the prime contractor often uses in his own operations. Absorbing a moderate per cent of defectives may be less costly in the long run than a rigid policy on rejections.

In what ways, if any, did the companies find currently available statistical quality control techniques inadequate for use in receiving inspection? This will influence the extent to which we can promote the techniques with vendors.

		Ranking Frequency 1=most important objection, etc.			
		1st	2nd	3rd	4th
(a)	Too complicated, difficult to teach inspectors.	3	4	4	5
(b)	Useful only on long runs of a fixed design, not on short runs.	16	6	2	
(c)	Too many risks involved for use on critical requirements.	6	11	4	
(d)	Quality can be adequately controlled for less cost through spot checks by experienced inspectors.	5		3	6

FIGURE III

Objection (b) crops up most often and most importantly. Category (c) might have bulked larger except that several companies took the stand that statistical sampling is forbidden anyway on critical requirements. This is true in many specifications and regulations. Nevertheless, there are some military production programs on which critical requirements are handled by a very strict sampling plan.

One or two companies have remarked that 100% inspection does not necessarily mean 100% conformance. This has led to experimenting with variables acceptance techniques that try to keep the critical characteristic a safe margin away from the failure point. The Lot Plot plan can be used for this purpose and the Lot Template technique is promising.

Ten companies gave statistical quality control techniques a vote of confidence. They felt that none of the objections were serious if planning and training were adequate. Others who listed objections would probably agree with the statement made by one, "Normally we find existing Statistical Quality Control techniques adequate for use in Receiving Inspection. However, like most techniques used in Inspection they are subject to improvement. ..... We wish to stress that we do not feel that the deficiencies noted make the techniques inadequate for use but are merely shortcomings which could stand improvement". Several also objected to spot checks as a substitute for statistical sampling as typified by the comment, "In general we find that we obtain better inspection which in the end results in less cost by using statistical control in lieu of spot checks".

An example of a well developed effort to put across sampling methods is illustrated in Figure III. The Bendix Radio Division has hundreds of suppliers to deal with. They have found it worthwhile to set up displays, distribute pamphlets, hold meetings and generally advertise the "how" and "why" of what they do in receiving inspection.

We have been discussing the improvement of vendor quality performance. This assumes that the company has a fairly exact idea of how well the vendors are doing. The questionnaire asked about the means Quality Control used in reporting vendor performance. Is the report made to purchasing? To general management? Four of the fifty companies do no reporting and nine pass only routine information on to Purchasing. The majority make some kind of summary report for use by management as well as Purchasing. Some use per cent of parts rejected and others per cent of dollar value rejected. The range of reporting techniques is indicated by the following:

1. "Quarterly report from Accounting Department listing dollar value of shipments received by vendor plus dollar value of rejections, rework and scrap. These figures are worked out percentage wise by Quality Control and control cards kept on each vendor and supplier. Report made to Purchasing and Management."

FIVE STAR ITEMS (Continued)

			Number of Pieces								
			Received		Rejected			% Rejected			$ Value Rej.
						Yr. to Date			Yr. to Date		
Part Number	Part Name	Vendor	Week	Yr. to Date	Week	HAC	Vendor	Week	HAC	Vendor	Yr. to Date
41-70-0687	Transformer	Whelan Elec. Mfg.	--	320	--	0	0	----	0.00	0.00	0.00
"	"	Standard Engrs.	--	100	--	0	88	----	0.00	88.00	990.00
"	"	Ace Elec. Mfg.	76	121	76	0	76	100.00	0.00	62.80	855.00
41-70-0981	Transformer	Standard Engrs.	189	189	0	0	0	0.00	0.00	0.00	0.00
"	"	Acme Mfg. Co.	203	217	66	0	67	32.51	0.00	30.87	1,326.60
43-27-100	Ver. Gyro	Steel Mfg.	14	68	2	0	5	14.28	0.00	7.35	6,375.00
560003	Accelerometer	Indian Scientific	10	182	10	0	25	100.00	0.00	13.73	7,387.00
560005	Control Tyro Roll & Pitch	Amos Elect.	11	86	1	1	12	9.09	1.16	13.95	13,463.06

Missile Manufacturing Division
Hughes Aircraft Company

WEEKLY INSPECTION TEST REPORT

Disposition of Suspensions Code

First Digit
1 Vendor Responsibility
2 HAC Responsibility

Second Digit
1 Vendor Pays
2 HAC Pays
3 Use as is

All Values in Dollars

For Week Ending 2-28-54

Vendor Number	Vendor Name	I. T. R. Number	Description	Total Value of Suspensions	Disposition of Suspensions	Prod. Class
				*		
580	A O M CAST	12446	116590 3	40	13	3
				40 *		
4120	AIRPAX	12473	992202 1	75	11	3
4120	AIRPAX	12474	992202 1	300	13	3
				375 *		
14760	C B S STEEL	12470	129281	712	13	3
14760	C B S STEEL	12492	129281	62	13	3
				774 *		

FIGURE IV - TWO EXAMPLES OF REPORTS TO PURCHASING

2. "Issue monthly a vendor rating list based on performance of last 30 days or ten lots."

3. "Monthly report of fraction defective in various categories of material is made by Receiving Inspection to Manager of Quality Control. This information is in turn incorporated in report to general management, manufacturing, purchasing, etc."

4. "Vendor quality performance on an individual basis is given periodic review in Receiving Inspection from vendor performance records and action taken where deemed necessary."

5. "A weekly report is made to Purchasing on vendor rejections. Further, a set of Statistical Quality Control charts on commodity rejection by quantity, type of error, etc., is provided Purchasing for buying group coordination."

6. "Report is made to both,
   (1) Graph by weeks - % of material referred to MRB
   (2) Graph by weeks - % of material reworked
   (3) IBM runs - results of each lot inspected by vendor and part number including process average, size of lot, reason for rejection, etc.
   (4) Special reports on key vendors showing quality performance."

Two examples of periodic reporting on vendor performance are shown in Figure IV. The cost information contained is also used to make separate reports showing favorable or unfavorable trends.

It appears that Quality Control in many companies has been doing some thinking about how to improve vendor quality performance. The tendency is to work more closely with Purchasing and Engineering. A sub-contract that starts out with a good understanding between these three departments has a better chance of coming out well. The modern techniques of Quality Control are getting a work-out. Some of the techniques are statistical, but many involve only a better organized approach to the question of "what is wanted and how do we go about getting it?" We have developed methods that increase the efficiency of our trouble-shooting, but we are even more interested in how to prevent the trouble in the first place - or to catch it before it grows into a disaster. Most important, the effort is being made to bring the vendor into the act; to agree with him on the rules of the game; to keep him out of trouble; to recognize that his success is important to our own prosperity.

BIBLIOGRAPHY

1. Howard T. Lewis, "Coordinating Procurement: A Case Study from the Aircraft Industry", Harvard Business Review, September-October 1953.

2. John A. Cairns, "Today It's Important to Cultivate Suppliers as well as Customers", reprinted from Printers' Ink, January 30, 1953, Copyright 1953 by Printers' Ink Pub. Co., Inc., 205 East 42nd Street, New York 17, N.Y.

3. "Quality Control in Component Manufacturing", reprinted from Tele-Tech & Electronic Industries, April 1953. Copyright Caldwell-Clements, Inc., 480 Lexington Ave., New York 17, N.Y.

4. Roscoe Smith, "Quality Level Certification", Tooling and Production, 6th Annual Quality Control Feature Issue, October 1953.

QUALITY CONTROL AT WARNER GEAR

Francis E. Jolliffe
Warner Gear Division

We are very fortunate in our plant that our work is such that statistical Quality Control practices can very easily be adapted to our manufacturing processes. We are a gear manufacturing shop. All of our operations are metal removing operations. We build transmissions, conventional, overdrive, and automatic, for automobiles, truck transmissions and transfer units for about twenty-five or thirty passenger car and truck manufacturers, as well as for many other manufacturers for miscellaneous small production applications. We have a capacity of more than 9000 transmissions per day, and employ over 6000 people. We purchase rough gray iron and die castings and steel forgings, as well as many semi-finished and finished parts, and minor assemblies, all of which are assembled together in our own plant and tested before being sent on to our customers.

In our organization the Quality and Inspection Divisions are all directly responsible to the Supervisor of Inspection, who is responsible directly to the President and General Manager of the company.

Our beginnings in the use of Statistical Quality Control were quite simple, and perhaps unorthodox. It was after our Chief Inspector and the speaker had attended a Quality Control School and visited in our customers' plants who were using Quality Control, that we made some data sheets which we gave to our inspectors in the Grinding Department and asked them to submit reports on samples of five pieces taken at random from their inspection benches. The Grinding Department was chosen as the guinea pig in our case because we had been having some complaints from our Assembly Department that the parts did not go together properly, that is, the bearing diameters would be oversize, causing the assemblers extra work to drive the bearings in place; or the holes would be undersize, causing the parts to go together too tight Also, grinding was a final operation, the work being performed just before assembly and shipping to our customer, and we felt if we were to improve our quality it should be in a position where our customers would receive the benefit of it as early as possible Another item, the supervisor in that department was very progressive and open to any and all suggestions that might help to reduce their cost. These inspectors data sheets were sent to the speaker's office where they were posted and kept in our office. This was done merely so we could gain experience in the making of Bar X and R Charts. After these preliminary charts were made on possibly a half dozen parts, they were shown to the Plant Superintendent and the General Foreman, who thought it might be a good thing, and were very willing to go ahead and help in the experimentation.

After we had the approval of the Plant Supervision we then took the charts and our plans to our General Manager

and our President and asked for permission to spend a little money for necessary forms and charts and to make some trial installations. Their approval was readily given and we have, at all times, had the fullest co-operation and approval of our Management.

We decided that our Bar X and R charts would be made in such a way that the charts could be kept in the department in full view of the operators and that the parts would be checked and the charts posted by our floor inspectors. Spaces were provided at the bottom of the chart for the inspector to show the sizes that he had obtained on each of the sample parts, and the average he calculated and the range that he found.

We use the low blueprint specification limit for 0 on the charts, making all of our samples read plus, unless they should be undersize. We believe this reduces the possibility of errors on the part of the inspector and helps him to make his plottings more accurately and in less time.

Our average and range values are plotted with large enough dots that they can readily be seen by the operators and the supervision when passing through the aisles of the department. We also provide a place for the inspector's initial, the time and date of the plotting, and a place to show what particular part was checked at that time, if the chart covers several different part numbers with the same size. We felt we could gain the confidence of the operators and supervisors if this were all kept open for their inspection, and also if they thought the inspector had erred in his checking they would know exactly what he found and could question it at the time.

The first chart placed in the plant was on a honing operation in the Grinding Department, honing the hole in our overdrive pinion. It was not a difficult operation and we had an air gage at these machines for checking the size of the hole. When the charts were made ready they were explained to the foremen of both the Production and Inspection Departments, and they in turn explained the project to their group leaders and operators and inspectors. Then came the big day when we put the charts at the machines and actually started our checking and plotting. This extra checking and plotting meant extra work by the floor inspectors, as we did not remove the bench inspection at this time. However, after the charts had been up for a few days the bench inspectors began to find less and less defectives, and of course, reduced inspection.

Our next step was to install charts on an entire manufacturing line. This was placed on our overdrive output shaft line, where it was necessary to grind the clutch teeth to length, internal grind the raceway for the free wheel rollers, two bearing diameters, and the spline diameter. Our rejection rate on this part had been quite high. That is, our salvage rate had been quite high, and we also had

considerable scrap and re-operation. The scrap ran into a great deal of money because it was an expensive part to manufacture and these operations were the end of the manufacturing process. Again, after the charts were made and explained to the foreman, he in turn explained the setup to the group leader of the line, telling him that it was purely an experiment and that it might or might not last. The group leader expressed his willingness to co-operate, so the charts were put on the machines. We did not talk to our Union Committee about the installation because, in the first place, we did not feel we knew enough about it to be able to explain it to them thoroughly, and also, since we felt it was still in the experimental stage, we would be better able to explain its workings after we had enough posting to know where we stood. This particular line worked three shifts, it had three pairs of internal grinders, that is, three operators operating six machines on each shift, a total of nine, all competing with each other to see who could make the best looking charts. I should mention that we use a color scheme to indicate the various shifts, using blue dots for the first, or midnight shift, black dots for the day shift, and red dots for the evening shift. Again, the floor inspector did the checking and posting, and the bench inspectors were retained. It is necessary to keep some final inspection to watch visually for missed operations or portions of the parts that do not clean up in the grind, and they also count and bill out the material after it is completed. Our original charts had the blueprint specifications indicated, but we had no control limits. However, we soon added control limits, which again required explanation to the operators as they felt at first that we were cutting down their tolerances. However, they soon learned that such was not the case, and they continued to co-operate whole heartedly, and did make a considerable improvement in the quality of the part.

Records kept on this operation for the first several months of its installation showed that our scrap was reduced from approximately 1% to about 1/10 of 1%, and the re-operation was reduced from about 6% to less than .5%. Holding the length of the clutch teeth had always been a considerable problem, and we had found it necessary to specify different thicknesses of snap rings for use in the assembly for a selective fit of the mating parts. After this operation came into control we were able to reduce the required number of snap rings to one. We learned, too, that internal grinders do not always grind to the low limit, and external grinders to the high limit, as most of us had been taught was the case. We, of course, had to replace our go and no go plug gages for the internal diameter with an air gage in order to have accurate direct readings that could be posted on the chart. We found very soon after the charts were installed that the grinding machines received considerable attention on repairs, and a couple of them were even re-built.

Our next steps were somewhat a surprise. Soon rather a tough group leader, making similar parts on an adjacent line,

requested charts. His group was producing good work, with less scrap and re-operation than the first line, but he still wanted his group to do the best work of any group in the department. The group leader was an internal grinder and very seldom produced any scrap. However, his own chart showed that while, from the beginning, he was producing parts within control, that he soon was grinding the raceways so close that the dots went right down the middle of the chart, proving that Quality Control helps good operators do even better. Then an operator was transferred from the first line to another line on the other side of the department. He soon asked the foreman for a road map for his new line, since, he said, the second line really needed the charts more than the first line. So we installed charts on that line for him. From then on, we began to get requests for charts on troublesome operations from the foremen and continued in that grinding department until we had a very complete installation. We gradually, as the quality improved, transferred our bench inspectors to floor inspection and eliminated the bench inspection, except for visual inspection and counting and billing. This guinea pig, or educational department, is still our pride and joy. The foremen in this department have sold Quality Control in a better way, and to more departments, than we in the Quality Control Department. We still call on them when we have a job that needs a little encouragement or extra selling. We did not find it necessary to change any blueprint tolerances for them, but rather, that they could grind the parts much better than they thought they could. However, we did find it necessary to improve some of the roughing operations ahead of them so that the material coming to them would be in better condition for their operation. To illustrate the point that they could grind the parts better than they thought, we have a free wheeling cam in our overdrive units that is very particular. The flat surfaces must be parallel with the axis within .0002" and square with the center line within .0002". The grinding supervisors wanted a chart put on there to improve the cams. However, they warned us to start that we would never be able to produce cams to the specified tolerances. Well, we didn't for a long while. But they finally kept working at the operation and getting the bugs out until they are now, and have been for sometime, producing cams that are almost 100% to the specifications.

After the Grinding Department we next tried our charting in the Blanking and Cutting Departments. We broke the ice on the green grind operation on the stem of the same overdrive output shafts that we started with in the Grinding Department. These diameters were green ground because they were used for locating points for cutting the teeth. Now, if there is anything that a gear cutter does not like, it is to have parts that don't go into his hobbing machine arbors easily. So the green grinders just hold their size down to the low limit, or below, to help the gear cutting operation. This is quite natural since they were all members of the same incentive group, and what helped one operation to make money helped them all. Naturally, if the locating points were undersize the shafts would have excessive

runout. It wasn't difficult to get these operations up to the blueprint specifications, since they were so nearly like the operations that had already been successfully charted in the Grinding Department.

The next operation charted in the Blanking Department was indicating runout on the end of the overdrive sun gear. This was important as the gear cutters used this surface for clamping, and if it were not square with the hole we would have trouble with runout and lead error, which of course contributed to overdrive noise. After a little work of correcting the broaching fixtures and improving the blanks coming to the broach, this operation was very much improved.

Then we tackled the problem of hobbing the flats on the overdrive cams that were mentioned before. These were a constant source of trouble and a great deal of argument between the hobbing operation and the Grinding Department, each contending that the scrap losses were the fault of the other. This operation took a great deal of work, but after repairing the arbors and the machines, and improving the blanks, and improving the gage, we got them to run fairly consistently, but could not hold the very close tolerances that were specified on the blueprint. So the tolerances were increased, and now they are being held almost to the original tolerances.

Another problem was in the hobbing of our spline shafts, the shafts on which the sliding gears of the transmission operate to select for the various speeds. These shafts have, for the most part, helical splines, and if they are off lead or if the keys are eccentric with the outside diameter, of course the gears won't fit properly. They either will not assemble at all, or if they do they will be hard to shift. Almost everybody agrees gear cutting is an art rather than a science. Our gear engineer not only checks the gears produced, but he also consults with the product engineers on new designs and specifies tolerances and he recommends equipment for the manufacturing and checking of the gears. Also, thetypes of hobs and cutters and shavers to be used. He is a very aggressive and progressive person and co-operated in every way in working out ways and means of checking and posting our findings. These splines really gave us a hard fight, but we did win. We are now using Bar X and R charts on the semi-finishing of the gear teeth since we know that our final shaving operations cannot make enough correction to make a good gear out of a poor semi-finished gear. In some instances we have revised our tolerances and our checking tools, and all concerned are very happy with the results being obtained.

On these gear cutting operations we cannot eliminate our final bench inspection, due to the work that is done after the operations being charted, so there is no saving on inspection. Therefore, we have used charts on some jobs to locate bad spots, and then have removed the charts after the machines were repaired or exchanged, or the tools or rough stock improved.

We decided to use Mil-Std. 105A as a basis for our receiving inspection and sampling, because we anticipated some work for the armed services, and believed that our regular parts could be used as a training ground in the use of Mil-Std. 105A. We sent our Quality Control and Receiving Inspection Supervisor to see one of our customers, who we believed had a very good sampling plan in operation. From what they learned we wrote a Manual for the instruction of both the supervisors and inspectors in this Receiving Inspection Department.

Our first objective was to gain information on quality levels of various parts that we were purchasing. Our Purchasing Department had sent out inquiries to most of our vendors to see if they were using Quality Control, and we got some very funny answers. Some admitted freely they had always used Quality Control, others that the price of the parts would be increased if we insisted on Quality Control, but most were somewhat evasive in their answer. In reply to our inquiry of what they considered a fair quality level, their opinions varied considerably. On springs, I remember that it varied from 5% to 20% defective. However, they all agreed that this percent defective was never mentioned in their quotations to the Purchasing Department, and the buyers were all laboring under the mistaken belief that they were buying 100% good material.

We first issued instruction sheets and began sampling on some ten or twelve oil seal rings and thrust washers, and it has gradually been increased until we now cover some 300 parts, including springs, stampings, die castings, snap rings, thrust washers, minor assemblies such as pumps, governors, brake bands, etc. We are confident that the data gathered on these sampling inspections is much more reliable than the former spot checks. This is due, we believe, to the fact that the inspectors have instruction sheets explaining just what they are to check and the tools or gages required for that inspection. Then they show on an incoming material record sheet the actual record of what they find in the individual shipments. This gives us a growing record of the performance of that vendor on that part. These findings are recorded in our Quality Control office on trend charts, on which are plotted progressively the average of the last ten shipments, and shows whether a part is becoming gradually better or worse, or whether they have erratic quality levels. The results are reported to our vendors through our Material Department and reports are also given to our Purchasing Department, to keep them informed of our findings.

Some parts are received from two or three different vendors and when questionable parts are received from one of the vendors we check the records of the others. We then contact the source of the questionable material. If the vendor being questioned thinks that that lot of material was as good as could be expected, we then refer to former lots of material received from them where it had been better, and also to the quality levels of their competitors. Generally, the next lots of material received from those fellows have been im-

proved.

We have not attempted to install a Certification program. We have not set any quality levels. We have just tried to improve our levels, calling bad lots to our vendors' attention and we found that we received better co-operation from them after we had shown them our sampling methods than we did prior to that time with our old spot checking method.

There has been a considerable improvement shown. The quality level on the first twenty-five shipments of ten given parts showed that of one and one-half million parts received, it was necessary to check 33,000 parts, which is $2\frac{1}{4}\%$, with a quality level of 21%. A check on the last twenty-five shipments of the same parts showed that, of more than two million parts received it was necessary to check only 23,000, or 1%, and that we found a quality level of 5%. This meant that we reduced our inspection from $2\frac{1}{4}\%$ on the first twenty-five shipments to 1% on the last twenty-five, less than half, with a resulting decrease in our inspection costs, as well as a considerable improvement of the quality level of the parts received. Now, we know that the Acceptance Sampling did not improve the parts, but we do feel that the information gathered on our sampling plan was of help to our vendors in making the improvement.

We find that we are spending less time in sorting the parts in our stock room. We have less delays at the assembly due to rejections and sorting of defective parts that were not found until they were ready to be used in the assembly. We have less complaints from our vendors due to rejections of parts that have been laying in our stock room for several weeks or months because defects are found when they are received. We have reduced our Test Stand rejections, and we have issued less orders for re-working purchased parts in our plant to maintain our assembly schedule. This has been done with no price increases due to the installation of our Quality Control program. We find that our acceptance sampling works very smoothly. Our principal difficulty is getting our own Inspection Department to do their work properly. On some materials that are difficult to pick random samples we find that the inspectors must be checked to be sure that this is done properly, realizing that if it is not done properly we are wasting whatever time is put in on the plan.

In our Assembly Department, we are, at the present time, using only one chart. All of our units are tested for sound and functioning 100% on Test Stands that simulate car conditions and then passed to a backline inspector who checks for wrong or missing parts, the torque of bolts, etc., giving us one last chance to look the job over before it is shipped to our customers. At one time we tried some "p" charts for our Test Stand rejections, but these did not work out because many of the causes for failure on the Test Stand were not due to assembly work. Also, the assemblers work as a group, and we find that one group will not compete with another group like one individual will try to do better work

than another individual. We found, also, that the foremen spent more time pointing to items that were not their fault than correcting the items that were their fault. At present we have one "c" chart in one of our Assembly Departments. It is based on purely assembly defects. It is based on the defects per 100 transmissions covering parts omitted or assembled improperly, screws improperly torqued, wrong parts, and so on. This has worked out quite well, we believe, because it is recording the work of only a small group and the supervisors have taken an active interest in trying to keep the plottings at a low level. This chart has shown an improvement from 15 or 20 defects per 100 units to 4 or 5 per 100 units, and it is not nearly as erratic as when first put in the department. This varies, of course, with changes of manpower due to schedule changes.

We have talked so far about our successful applications. We have had a few discouraging ones. Generally we found that our failures were due either to improper charts or to the men or machines not being made ready for charts, or we hadn't properly acquainted the men with their use and interpretation. Some machines are not capable of producing parts to blueprint tolerances and work much of the time to salvage limits. To re-tool the entire line was too much of an expense to ask for, and so charts on these operations have been withdrawn. Many changes have been made on the operations on which we have used charts, such as having the machines repaired or machines exchanged, gages changed from go - no go to direct reading air or dial indicator or electronic gages. With reference to the machines, our Grinding Department superintendent called at the plant of one of our grinding machine suppliers, taking with him some of the Bar X and R charts that had been used on a machine in question. After explaining the charts and our method of gathering the information, they were very much interested and asked if they might not have the charts for reference. Fewer engineering changes on our blueprints are being made. We find that not many are needed. Accurate checks and records have shown that most pieces are well within our present tolerances. That record is shown to the manufacturing group whenever a job does go bad and they work toward getting the job back on the beam instead of requesting more tolerances.

In an attempt to stimulate the interest of the production foremen in the charts in their departments, and encourage them to use the information shown on them, we borrowed a plan from one of our friends that has been a great help. A large board was made and placed in the Superintendent's office containing the part numbers of various parts over the plant that were giving trouble, and on which the charts were showing out of control. The part numbers were listed together with a list of the charts of the operations on that particular part, and with a hook beside each chart number that could bear a tag that was green on one side and red on the other Whenever the inspector in the department found a certain chart out of control he would phone the superintendent's secretary, giving her the number of the chart that was out of control and she would turn the tag over to show

red instead of green. As soon as the superintendent would come in and notice that one of the charts was red he would immediately call the foreman of that department, asking him the cause for the difficulty and what had been done to straighten it out, and whether it was going properly at that time. At first, this, of course, caused a little embarassment on the part of the foreman, since he had not noticed that the chart was out of control. But they soon got the idea, and now watch the charts more closely and generally, when the superintendent calls them, they already have the answers ready for him. Of course, when the chart is again in control, the inspector calls the secretary again, who turns the tag back to green. It is this close co-operation between our management people, the Purchasing Department and the Quality Control Department that has made our Quality Control Program the success that it is.

With regard to the operating personnel for our Quality Control charts, we have all dimensional checks and posting done by floor inspectors. This has been done with no overall increase in manpower. As different jobs were brought into control, the bench inspection work was reduced and those bench inspectors were put on the floor to operate charts. In fact, there has been a slight overall decrease in manpower. The exception to this has been, as mentioned before, in the gear inspection. The charts there are used on semi-finish operations and the bench inspectors are still required to check final operations. All inspection work is done under the regular gear or general inspection supervisors. Our Quality Control men do work with the inspectors on the floor, but only in an advisory way. All direct instruction or criticisms are handled through the regular inspection supervision.

In our Quality Control office we now have three men and one girl. They are all responsible directly to the Supervisor of Inspection. That is, they are entirely separated, as far as responsibility, from the Inspection Department. These men investigate jobs that are not going properly, and collect data, design charts, install the charts and instruct inspectors and supervisors in their use in certain areas of the plant. They also follow up the posting of the charts in the shop. They examine all completed Quality Control charts that are removed from the machines and sent to their office, and make reports of any charts that are improperly posted to the inspection supervision, and on any charts that are out of control to the manufacturing foremen and superintendents. And, by the way, we investigate charts that look too good, as well as the ones that look badly out of control. This investigation of extra good charts is based on the possibility that there may be something wrong with the way the information is gathered and posted, or we may learn from an exceptionally good chart what has been done to bring the job into such excellent control and that information may be of great value to us the next time something goes wrong. These Quality Control men also make capability checks on questionable operations on machines, generally at the request of the manufacturing supervision.

The girl in the Quality Control office does all the stenographic work, and calculates the grand average for averages and ranges on charts that are out of control or erratic, and posts these grand averages on what we term progress charts, which gives us a running history of the performance of this particular job over many months. Blueprints of these progress charts are sent to the foremen and superintendents of the departments involved.

We now have about 475 different Bar X and R master charts covering approximately 2000 different parts or operations. The masters for these charts are made on specially printed vellum and the heading consists of the part number, part name, operation name, etc., and the blueprint sizes and units of measurement are all typed in the margin. Black and white prints are made from these masters and a supply is kept in the inspection desks in the Manufacturing Departments. The floor inspector replaces charts on the boards on which they are mounted in the department when they become full, and requisitions more black and white prints when his stock gets down to a minimum number. We have some universal charts; for example, those on which runouts in the gears are plotted, which are purchased, printed on white paper ready to mount on the board in the department. These are issued to the inspectors as they are required.

I wish I knew how much we have saved in dollars, Warner Gear that is, not our Quality Control Department. We realize, and know, that any savings made have been the result of the direct action taken by the Manufacturing Department and that our charts are merely a means of giving them information as to where trouble might lay, and how extensive it is. Our Cost Department records do not show the cost of scrap and re-operation on individual parts, but information taken from our daily inspection reports shows a decrease in the number of pieces scrapped and the number of pieces re-operated on parts on which charts have been installed. Also, by comparison of production charts with charts made from data that was collected before the installation of the charts, we know that the operations have been improved. However, even if there has been no reduction in cost, the reduction in complaints from our assembly, the reduction in our Test Stand rejections, and our customer complaints and in the number of transmissions returned as defective, shows that our program has been very worthwhile.

There has been an increasing interest on the part of all employees; our foremen, because of seeing various jobs improved by getting data to back up their claims of need for better tools or equipment, and by meeting their production schedules easier because of having less scrap and re-operation to use up their machine time. We recently checked with some foremen on the possibility of removing charts from jobs that have been in control and giving no trouble at all for sometime. They were unanimous in their belief that the cost of maintaining these charts, even though constantly in control, was worthwhile because they can know by glancing at the chart that the operation is still going satisfactor-

ily, or if it does start to get out of hand they know it immediately.

Increasing interest on the part of our Processing and Tool Department is shown by the fact that they ask us for data on certain jobs. This in connection with consideration of tooling new jobs and considering new methods, as well as on improving present ones.

Although we have never gotten around to a general discussion of our Quality Control plans with the Union committee, it has come up occasionally in meetings, but never with anything derogatory. In fact, they have expressed the belief that now the Company people really have knowledge of wh they are talking about where before they could not have been so sure of their ground. We recently investigated a report that one of our Union leaders had complained about the Quality Control setup, but talking with him we found that his gripe was with the earlier operations rather than on his own. It was his idea that we should put charts on more of our earlier operations so that he would have more uniform parts coming into his department, thus reducing his trouble.

In schools sponsored both by the Purdue Extension Department and our local ASQC Chapter, there has been a yearly increase in attendance. Warner Gear encourages our people to attend these schools. This year we had 47 Warner Gear employees enrolled. The company pays the tuition for foremen and inspectors, but the people attend the schools on their own time, attending 2 hours a night, one night per week, for ten weeks to get their basic training.

Our superintendents and works manager are always willing to co-operate and to discuss our problems and the results of our investigation, and initiate or expedite remedial action wherever possible.

Our Purchasing Department appreciates information that we can and do give them with regards to various parts or vendors, and are always willing to call in vendors for discussions. Our Sales Department, in their contacts with our customers, point to our Quality Control program as a means of guaranteeing the continuance of the good quality on which Warner Gear has been built.

In spite of the fact that we still slip occasionally and get bad work, and charts get out of control, and that we then must dig, dig, dig, all over again, I have found no one in our Company yet who is ready to lay down the program, but rather we are constantly increasing our Statistical Quality Control operations.

OPERATIONS RESEARCH CONCEPTS USEFUL IN QUALITY CONTROL

Harold O. Davidson
Operations Research Office
The Johns Hopkins University

The title of this discussion might be regarded either as a statement which is asserted to be true, or as a proposition which may or may not be true. I prefer to regard it as a proposition which may or may not be true depending upon what you mean by "Quality Control" and what I mean by "Operations Research." In a discussion of this sort there is some advantage in starting at a point where nearly everyone can be in agreement. To this end I think it can be stated that Quality Control and Operations Research are broadly similar activities in these respects:

(1) They are directed to (materially) useful objectives;
(2) They deal with types of problems, which can be analytically generalized, in working toward the objectives;
(3) They have developed or adapted objective techniques for solving these problems.

This is perhaps an overly simple way to characterise a general class of activities of which both Operations Research and Quality Control are members. However, this scheme does serve to indicate the relative order of importance and consideration: objectives, problems, techniques. Without a clear definition of the objectives one cannot intelligently formulate a problem, and without a formulation of the problem it is rather useless to grope around for techniques.

I would like to emphasize as forcibly as possible this order of consideration - objectives, problems, techniques. As I will show later it is the order which Dr. Walter A. Shewhart followed in his pioneering quality control investigations. There is a strong tendency, however, for many persons both within and without a given profession to define the particular field in terms of its techniques. This historical approach does have a certain pragmatic appeal, for by stating rather specifically what the capability of a profession has been in the immediate past it indicates approximately what it can do in the immediate future. There is also a pronounced tendency to concentrate effort on further exploitation of techniques which have proven successful while neglecting the problems for which adequate techniques have not yet been developed. The complacent indulgence of this tendency is a sort of professional malady which I would call technique-orientation. A strongly technique-oriented profession will continue to improve its traditional methods of solving problems but is relatively sterile insofar as the innovation of new techniques is concerned. The mere improvement of traditional techniques, however, will rarely avail to keep a profession abreast of the dynamic industrial society in which the important problems of one generation are not necessarily the important problems of the next. The technique-oriented profession, then, may show a continued improvement of techniques, for the solution of decreasingly important problems - because it has lost sight of the original objectives upon which the profession was founded or because the original objectives are no longer the proper ones.

Every profession is susceptible to this disease and suffers from it to a greater or lesser degree. It has been very apparent in the time

study area of Industrial Engineering. It has appeared in Quality Control. And I have observed signs of it in as new a profession as Operations Research. In speaking of technique-orientation as a disease I intend a quite literal analogy. It is definitely contagious, being transmittable not only by persons who are infected with it but also by technical journals. It saps the vitality of a profession; and it impairs the profession's usefulness to society by virtue of an accompanying disability - to which Operations Research people have applied the term "sub-optimization."

Sub-optimization results whenever a problem, or part of a problem, is defined without adequate reference to its context; so that when an apparently best solution is found, it is not in reality the optimum solution with respect to the over-all problem. In many cases this amounts to solving the wrong problem, or to the solving of trivial problems. The sub-optimization fallacy obviously is not something that was just discovered by Operations Research. I am quite sure that it has been noted by people in every profession. At the same time, the general reaction in most technical professions is to conclude that this is a matter for somebody else to worry about; "it's outside <u>our</u> field."

The inadequacy of this viewpoint lies in the fact that problems in the real world are not broken down into neatly separable packages labeled Sales Problems, Quality Control Problems, Production Problems, Labor Problems, and so forth. These divisions are something that we impose rather arbitrarily, though with the admittedly sensible idea of reducing the general problems to manageable bits that people can work on. Troubles begin when each person runs off into his own departmental hole to work on his bit of the problem. And the troubles converge to torment the executive when the sales department informs him that the product must have longer life, better appearance, greater efficiency, and lower cost than the competitors' products. The production department insists that quality standards must be lowered in order to meet desired production schedules. Design engineering claims that the standards must be raised in order to achieve the specified wearing qualities of the product. Industrial Relations complains that excessive rejects are causing employee dissatisfaction with the wage incentive system. Meanwhile, the quality control director has been squeezed by the same pressures so that he, at least, may offer the managing executive something which the others probably do not; namely, sympathy. The technical work of quality control, however, is all too often as much a sub-optimization as any of the rest. I will offer illustrations in support of this point later in the discussion. At this point I would like to make sure that I have presented a reasonable view of the sub-optimization problem.

In speaking of persons working independently in their own departmental holes I am speaking figuratively with respect to professional methodologies. I do not mean to imply that no communications exist between persons, or that there is a total lack of mutual interest in achieving a genuine optimization of the general problems of the enterprise. Communications, both informally and by committees, do exist between departments. And not infrequently there is a substantial realization of the necessity for optimizing the solutions to problems on a higher-level than has been accomplished in the past. From my own observation I would say that his attitude is somewhat more common in progressive quality control organizations than in most other technical departments. In enterprises that are not too large competent people

working under these conditions are often able to achieve a very good degree of optimization for the general problems of the company. But as the size of the company increases the communications between departments tends to be less effective, and the complexity of the problems increases so that the determination of optimum solutions becomes virtually impossible without the aid of a specialized approach and methodology. The idea expressed here is no more radical than the notion that a civil engineer can design building structures more efficiently by means of his specialized approach and techniques than can a construction foreman by means of rough "principles" and guide rules. The Operations Research viewpoint holds that organizations and their operations are also "structured" and that it is possible by scientific investigation to discover useful knowledge about them. This provides the basis for an approach and methodology for efficiently determining optimum solutions for the organizational-operational problems. In brief, this is the central concept of Operations Research.

I am quite aware that many persons take a rather pessimistic view of the possibility of applying analytic methods to resolve diverse elements of large-scale problems and produce optimum system designs and operational procedures. Anyone who has engaged in the sort of Operations Research I am talking about would certainly a gree that the task is a forbidding one and that the methods available today are by no means everything that we would like them to be. Even so the accomplishments of the Operations Research profession during its first ten years are ample evidence for belief that its central concept is sound. This is not always clear to an observer who is puzzled to find that some work labeled Operations Research could in all probability have been done just as well by one of the other professions. The only remark which can be made to this point is that there have in fact been a dismaying number of instances where the Operations Research label has been indiscriminately or mistakenly applied. No doubt some of these are the result of genuine confusion; for by a technique-oriented inspection Operations Research looks like applied statistics from one aspect, Engineering like Industrial Engineering from another, and like Applied Physics from still another. And this is not surprising, since the founding of Operations Research as a profession essentially represents the gathering together from many different professions of persons who for some time had been working toward the same central concept. The amalgamation of techniques from many different sources was a quite natural result. This in itself has had obvious advantages. Perhaps an even greater advantage accruing to a coalition of professional backgrounds is the fact that it develops an improved facility for bringing the apparently diverse aspects of a problem into a consistent perspective. In other words, it supplies something which our compartmented and specialized system of education generally fails to develop.

What I have attempted so far is not a comprehensive definition of Operations Research but rather an exposition of the Operations Research viewpoint. I have indicated that it is a viewpoint which was developed simultaneously by people in many professions, and I am sure that it is to be found today in the advance elements of every profession. Nevertheless the great bulk of Quality Control practice today is an extension of the work of Dr. Walter A. Shewhart and is based to a great extent on the objectives, problems, and techniques as he defined them:

> ". . . we shall start with the aim of the engineer to manufacture a product of uniform quality. We shall take this

"to imply that the quality should be reproducible within limits, or that the engineer should be able to predict with minimum error, the percentage of the future product that will be turned out by a given process with a quality within specified limits. The engineer desires to reduce the variability in quality to an economic minimum. In other words, he wants

(a) a rational method of prediction that is subject to minimum error, and
(b) a means of minimising the variability in the quality of a given product at a given cost of production." (1)

This is the general point of view from which began the development of a considerable array of useful techniques applicable to problems of process control and acceptance sampling (and, incidentally, to problems of similar analytic form in many other fields). There is no need to ask whether this has been worthwhile. The important question now is of the future development of Quality Control. Which direction should it go? Should there be a change of emphasis to other aspects of the quality problem?

Whether Operations Research can make any substantial contribution to this inquiry is something for the Quality Control profession to decide. I can say, however, that quality problems arise in every OR investigation and that there is a general method of approaching such problems, from the Operations Research Viewpoint, which may be of interest to Quality Control people. As I have mentioned previously, the objective of Operations Research is the solving of problems in a way which will be optimum for the enterprise as a whole. This means that the goals of the enterprise must be clearly identified as a context for the problems considered. Without this step it is impossible in military operations, for example, to determine whether the purpose of a withdrawal action should be delay of the enemy (as in the Bataan campaign) or conservation of one's own force (as in the Dunkirk evacuation). I might mention in passing that those who have the time and interest will find a dramatic illustration of conflict between sub-optimization goals in historical accounts of the Dunkirk operation in World War II. This case is particularly interesting since continued insistence on a sub-optimization goal might well have been fatal for the enterprise as a whole. Briefly, the withdrawal was conducted with the major objective of conserving British forces. Sub-optimization of the Dunkirk operation clearly indicated a necessity of local air superiority. On the other hand, conservation of British defensive air power required denial of air cover for the operation. The decision to discontinue air cover was undoubtedly a painful one, but the narrowness of the margin by which the subsequent Battle of Britain was won indicates how close to disaster the British nation had been pushed by temporary sub-optimization of the Dunkirk operation.

While commercial enterprise does not offer such clear-cut or dramatic illustrations of the principle, it is unquestionably important to know from the outset whether the organization goal is, for example, short-run maximization of profits or long-run stabilization of profits and jobs. This has an obvious bearing on the quality problem, and is suggestive of the importance which should be attached to identification of the organization goals as a prerequisite to definition of specific operating problems. From the operations research viewpoint the identification of organization goals is essential to the formation of valid

"pay-off" functions.

The next step in an Operations Research approach is formulation of the problem. There are various ways of going about this but one which is often productive is to ask where the problem actually originates, and then by tracing its effects to determine the "system" with which we must be concerned in order to solve the problem. Studying the quality problem system is a sort of loop which begins and ends with the consumer. This conclusion is forced upon us I think by the fact that the ultimate disposition of commercial goods and services is in the fulfillment of some consumer need. In a fundamental and important sense the "product" of the entire loop process is the effect produced for the consumer (i.e., satisfaction of his requirements). Considered in this light the concept of quality is related to the Operations Research concept of effectiveness. Effectiveness is simply the extent to which a particular article or service satisfies the consumer's requirements. By fixing a certain set of requirements as a reference point we can assign fixed values of effectiveness to alternative articles or services and thus determine what we call quality. Effectiveness is therefore a somewhat more general concept than quality. For purposes of this discussion, however, the similarities are more important than the difference. Whether we use the concept of quality or the concept of effectiveness we are faced with the fact that there is no practical method or scale for directly measuring the degree of satisfaction of consumer requirements. In both Quality Control and Operations Research, therefore, we are forced to employ approximate measures of quality or effectiveness. In other words, since quality or effectiveness cannot be measured directly we are reduced to estimating it by means of criteria.

This problem can be illustrated quite simply by considering the quality of a Diesel engine, a common commercial article of manufacture which is capable of satisfying certain consumer requirements for power. An important quality characteristic of this article is its service life. It is unlikely, however, that the manufacturer will apply quality control techniques to service life. Instead, they are applied to such properties as dimensional variation, surface finish, and hardness of certain parts of the engine. This is done because there is reason to believe that these properties are functionally related to the quality characteristic which is really important to the consumer. The fact remains, however, that the control activities are governed by the criteria and not by the fundamental quality characteristic.

Criterion bias is one of the frequent and often unrecognized sources of sub-optimization, both in Quality Control and in Operations Research. One of the most interesting forms of criterion bias is scale distortion. A simple illustration of scale-distortion bias comes to mind in connection with the manufacture of booby-trap firing devices during World War II. Considerable quality control effort was expended with the objective of securing high firing reliability of the devices. This naturally had adverse effects on the cost and on the production rate. A quick analysis suggested that if quality standards were reduced there would be an increase in production and lowering of costs which would more than offset the losses due to the transportation of duds and the military effort expended in employing them. But this also was a distinctly sub-optimal solution, and resulted from a failure to get to the base of the quality problem. It had been assumed that the function of booby-traps was to produce enemy casualties and the military

effectiveness (or quality) of a dud was therefore zero. Consequently, the military effort expended in using a dud was erroneously considered a cost, in spite of the known fact that dummy booby traps were frequently employed. This latter observation was a clue to the fact that the primary military effect of booby traps is not the production of casualties but delay of an enemy advance produced by the necessity of locating and destroying booby traps which might inflict casualties. It became clear that from the standpoint of military effect the quality of the firing devices increased in some non-linear fashion, as the reliability increased up to a certain point beyond which there was virtually no gain in quality from an increase in reliability.

Another clear case of scale distortion arises in connection with the determination of quality of an infantryman's rifle. Insofar as the consumer is concerned the major product is enemy casualties, when the weapon is employed in defensive fire. The quality of a rifle in this type of action is its ability to produce casualties. Obviously, one of the important quality characteristics is accuracy of fire - not accuracy of the rifle, but accuracy of fire. This amounts to stating the self-evident fact that the military effect is produced by a man-rifle system, not by a rifle alone. It is apparent, then, that as the dispersion contributed by the rifle becomes appreciably smaller than that contributed by the rifleman its effect on the quality of the result becomes negligible. Hence the accuracy of a rifle, measured on the test range, is a scale-distorted criterion of quality.

I have chosen the preceding illustration because it brings in another matter of interest. The discussion of this problem appeals for an orientation toward end-results (products) of man-machine systems in operation. I think that there is implicit in this viewpoint a concept of "assignable causes" which is extremely useful. If we were to apply the quality control concept of assignable causes to the performance of the man-rifle combination we should probably study the dispersion of the individual shots with a view to detecting the effects of possible non-random causes. This might in fact be an interesting application of process control techniques, though I suspect that an experienced coach can detect most of the important non-random factors without this sort of assistance. Suppose now that we have a skilled rifleman and that the process is in a fairly good state of control from the Quality Control viewpoint. From the Operations Research viewpoint we will still infer the presence of "assignable causes" in the sense that the supposedly "homogenous" dispersion is in fact the joint effect of two independently acting cause systems. And from the Operations Research viewpoint the factoring out of the total effect the contributions assignable to elements of the system, whether independent or interacting, is essential to the determination of optimum solutions in the over-all sense.

I appreciate the fact that Dr. Shewhart's work was based on the tenet that one accepts the existing process and that in relation to this tenet the alternative interpretation of "assignable causes" is not particularly meaningful. It is also quite apparent that a great deal of useful work can be accomplished within the boundaries of this tenet, though at times it may be able to produce nothing better than a low order of sub-optimization by virtue of the fact that the solution to the quality problem is defective in some other part of the system. Perhaps what I am doing is inviting the Quality Control profession to take a bigger bite of the quality problem. Judging by past accomplishments I

would guess that the results of such action would be beneficial to industry.

This discussion has been suggestive rather than exhaustive of the community of interest, actual and potential, which exists between the fields of Quality Control and Operations Research. In addition to the analytical concepts which may have usefulness in both fields there are problems of determining costs - such as the cost of defective product - which are troublesome to both. These problems I have not touched upon at all for want of space. Finally, because of the need for diverse professional backgrounds in an Operations Research group and because of the particular contributions which the professional Quality Control engineer can make I predict that the community of interest between the two fields will be cemented by the addition of increasing numbers of Quality Control engineers to Operations Research teams.

(1) Shewhart, W.A., Statistical Method From the Viewpoint of Quality Control. The Graduate School, Department of Agriculture, Washington, 1939.

STATISTICAL QUALITY CONTROL IN A PETROLEUM REFINERY

John T. Walter
School of Business Administration
University of Pittsburgh

Statistical quality control in a petroleum refinery calls for a technique to control the continuous stream type of process. The principles involved also have application in large chemical plants and wherever there are extensive fluid operations.

An experiment was carried on in a large refinery over a period of several months, and the indications were that statistical control of a continuous fluid process can reduce the variability of product quality considerably. In these terms statistical quality control is a demonstrated success.

The project was initiated as a result of examining sampling schedules for a refining unit. These schedules were originally established to satisfy the demands of conservatively minded production men by calling for regular and frequent sampling. Attempts at economy in laboratory operation exerted a pull in the other direction, looking toward longer intervals between samples. In an effort to resolve this conflict and to establish acceptable standards of sampling frequency, applications of statistical quality control were investigated. It was assumed that a process operating in statistical control is a prerequisite to determining optimum sampling schedules.

The unit selected for experimentation was a fractionating tower with an average daily run of 62,000 barrels. The top stream coming off this tower turns out 20,000 to 23,000 barrels per day. This was the stream chosen. Product samples were drawn from the line at eight-hour intervals.

Ready-made procedures could not be found in the literature of quality control. At the most, hints were given that helped to point the way in the right direction. In fact, during the 1940's some academic and industrial statisticians were known to have said that statistical quality control had no place in the chemical industry. Owen L. Davies, in his Statistical Methods in Research and Production with Special Reference to the Chemical Industry, 1949, says of quality control, "Its application in the wider field of the chemical industry is still a matter of some doubt, but in view of its striking success in other fields, it is felt that the possibility of its application to chemical plants should receive careful consideration." As recently as February, 1953, L. J. McGovern, writing in the Petroleum Refiner, said that to his knowledge there have been no successful direct applications of statistical quality control to petroleum processing units. He said further, "The development of this approach is still on the drawing board." Thus, it can be seen that pioneering activities were necessary in the year of 1953. A start was made with only the assurance that no harm could result from the attempt.

Improved quality control presents other than technical problems. Some operating men, long accustomed to established procedures and having a conservative attitude toward their work, are slow to accept new ideas. It is only natural that people take pride in what they

have done for a long time. Unfamiliarity may also cause resistance to develop when new terms are introduced from the field of statistics.

The usual average and range charts require modifications when applied to a continuous stream. In metal-working industries five or ten samples at a time may be taken off the assembly line for inspection to determine their average quality. It is meaningless to average several samples drawn at one time from a fluid product stream. Fluids are homogeneous in nature, so each sample will be the same as another. Such averages would not be sampling averages. They would represent an averaging of laboratory test results, tending to eliminate measurement error. In this event, product quality is not being averaged by the statistician. Averaging becomes instead a chemical reaction.

To get a sampling average it is necessary to take the samples at different periods of time. If the intervals between samples are short, it comes close to being like the situation where two or more samples are drawn simultaneously. If they are farther apart, one is forced to wait some time after taking the first sample before final reports are forthcoming. This delay may be too long to satisfy operating requirements.

A chart of individuals gives quick results, but with certain other disadvantages. It lacks the smoothing effect of an average chart (see Charts 4 and 5 which present the same information in both forms). This smoothing out of short-term fluctuations permits more careful determination of the real process level. In addition, laboratory analytical error creates more uncertainty for any single sample than for an average of several samples. To say the same thing in another way, multi-testing increases the reliability of laboratory tests. Sampling error is affected in a similar way.

As a compromise between the chart of individuals and the chart of averages, a moving average was plotted. This seems to combine all the advantages of the average with some of the desired sensitivity to change. It even smooths out the curve more than an average does. While not quite so responsive to change as the chart of individuals, it is, perhaps, desirable not to be. On the other hand, the moving average does respond to each new test result incorporated into the average. Basically, the average chart is the guide to process level, while process variability can be better portrayed by a chart of ranges or standard deviations, a moving range of two samples being useful (see Charts 2 and 3).

How many samples should be included in the subgroup for the moving average? This is a debatable point. In the experiment under consideration a moving average of four was used. Averages of four are statistically more effective than averages of two or three. Distributions of averages of four or more tend to be normal, while the larger averages have less sensitivity to change.

Control limits are established in much the same way as in other kinds of quality control work. However, since there is no rational subgroup basis for computing ranges as a measure of process variability, the alternative is to form groups out of samples taken in consecutive order. From twenty-five to fifty groups are used. Factors from published tables of multipliers may be applied to the average range to get the desired control limits.

ILLUSTRATION

25		29		31		20		24	
22		27		24		26		25	
28		27		32		28		24	
22	6	27	2	27	8	27	8	18	7
22		26		26		26		26	
43		28		24		26		15	
46		29		24		28		21	
20	26	34	8	22	4	27	2	19	11
24		30		22		26		21	
35		24		26		26		20	
22		31		22		25		20	
28	13	25	7	22	4	26	1	20	1
22		24		28		25		23	
20		20		24		22		26	
20		25		22		20		28	
20	2	20	5	26	6	30	10	24	5
32		29		24		24		14	
20		26		20		24		17	
21		29		20		22		23	
24	12	28	3	20	4	23	2	14	9
24		25		24		22		14	
26		30		22		22		22	
24		28		24		24		16	
27	3	26	5	26	4	22	2	16	8
25		29		26		29		14	
26		29		23		26		18	
36		28		24		19		14	
27	11	36	8	27	4	26	10	23	9
26		32		28		18		19	
28		33		29		19		19	
27		26		24		26		16	
35	9	26	7	27	5	26	8	15	4
26		32		27		32		Total	329
27		24		20		20		Average	6.85
29		28		28		28		(2.282)(6.85)=15.6	
25	4	24	8	24	8	34	14		
23		28		24		20		Exclude	26
26		22		26		22		Total	303
26		22		23		28		Average	6.45
36	13	24	6	25	3	18	10	(2.282)(6.45)=14.7	

3 Sigma Control Limits for Averages (0.729)(6.45)=4.70

It is reasonable to look for some improvement when the chart is first installed, not knowing how much improvement can be made, of course. In the experimental project an arbitrary squeeze was put on the control limits by starting with 80% rather than 3 sigma limits. Success in staying within these limits proved the decision to be justified. Later 3 sigma limits were used, but by then these limits were narrower than the original 80% limits.

Final boiling point on a distillation test was the critical control. Prior to beginning statistical quality control the process had a standard deviation of 6.2 degrees over a period of fifteen weeks. Then, for a ten-week period a control chart of individuals was used

with a standard deviation of 5.0 resulting. Finally, a moving average control chart applied over a period of nine weeks produced a standard deviation of 3.1. The moving average reduced variability to one-half of what it had been without statistical quality control (see Chart 1). These gains were made without using a range chart, showing the effectiveness of a chart of individuals and then of a chart of the moving average uncomplicated by other considerations.

Reduction in variability of the product stream has several consequences. The immediate objective is to attain a desired cut point. This ultimately becomes a question of meeting customer specifications with greater accuracy, avoiding some degree of former quality giveaway. When applied to intermediate stages of production, statistical quality control means less difficulty with subsequent operations as a result of stabilizing their feed. In the fractionating tower itself adjoining streams will have less variability as a result of stabilizing the one stream.

In other industries statistical quality control has made it possible to abandon 100% inspection without any sacrifice of quality. There is no such thing as 100% inspection in a continuous fluid process, but the amount of inspections can be less with statistical quality control than without control. During the course of the experiment it was possible on two occasions to test more than sixty samples without making a single process adjustment. In the state of statistical control that this condition represents, fewer samples could have been taken without disadvantage. Twenty samples would have served as well as sixty, or one per day instead of one each shift. The laboratory reports that a reduction of two samples a day would mean a saving of about $3,000 per year in laboratory costs. Any such reduction is feasible only while the process is in control. Special samples are required more frequently during an upset.

It appears as though the only way to determine the right number of samples is to try different possible arrangements and judge by the results. This calls for some boldness on the part of management. Fear has been expressed that less frequent sampling may conceal an upset beyond the time when it would otherwise be discovered, with losses so great as to overshadow the savings from sample reduction. A cursory examination of some of the records raises the question as to whether distillation tests do not have their greatest usefulness in determining the process level, while other indicators serve to reveal upsets. Apart from unusual disturbances, it appears that control is not improved by frequency of sampling. In fact, the indications are that plant operators made too many process adjustments when not in statistical control, and actually improved control by adjusting less frequently (see Chart 6). This is one of the benefits of using the moving average chart. By smoothing out the curve the frequency of adjustment is reduced.

Statistical control of laboratory testing procedures is another application of the general technique (see Chart 7). Petroleum process control is dependent upon laboratory test results, and it follows that smaller laboratory analytical errors will mean improved process control. Periodic determination of testing error is a part of good laboratory management. Education of personnel or other steps toward improvement should be undertaken as soon as these investigations reveal where

problems exist. The best approach is to use statistical control charts in the laboratory. This helps bring the laboratory procedure into control and is the means of maintaining control with the minimum of effort. Complete success with plant control requires this supplementary effort in the testing laboratory.

It is important to know what to expect from quality control. Control charts are in no sense automatic in their operation. One cannot sit back and let the charts control the plant. Neither do the charts require much time and effort. The same amount of effort brings better results with statistical control than without it. The statistical approach to quality control can be doubly rewarding where there was formerly neglect. Control charts are merely better tools capable of superior workmanship. Either with or without their aid, skillful operators get better results than careless operators.

(Reprinted from Petroleum Refiner)

Chart 1 AVERAGES OF GROUPS OF FIVE RANGES OF FOUR SAMPLES

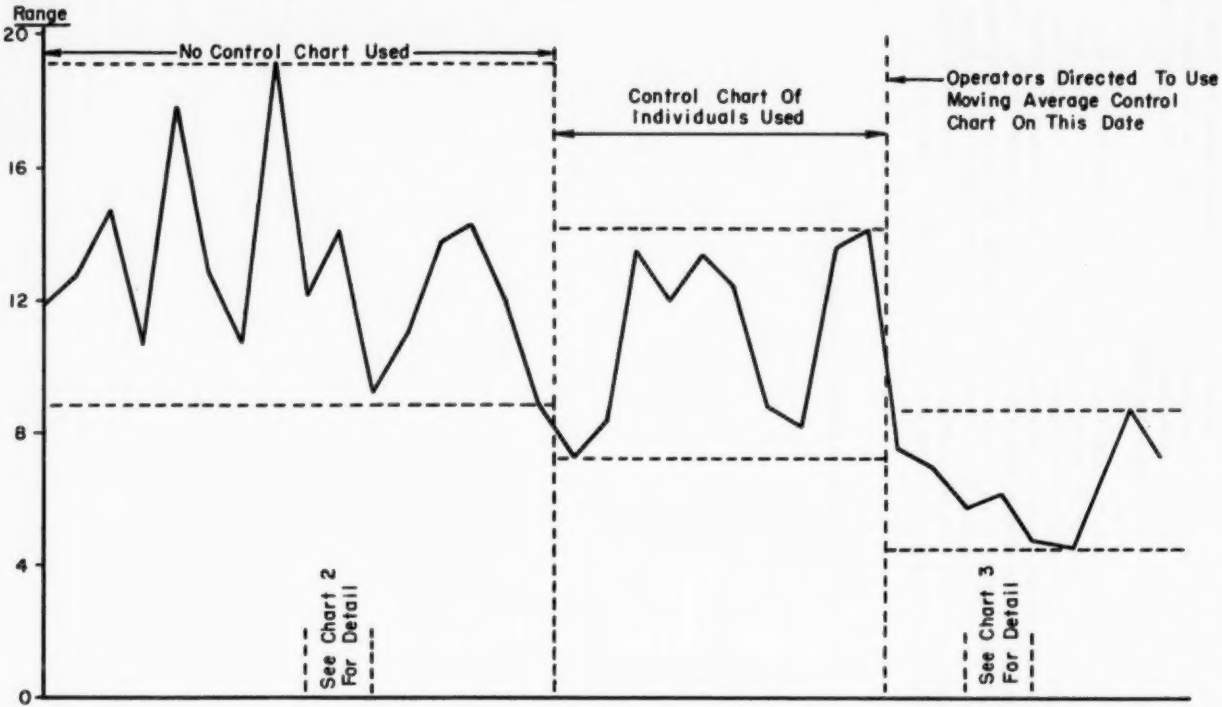


Chart 2

MOVING RANGE OF TWO
(No Control Chart Used)

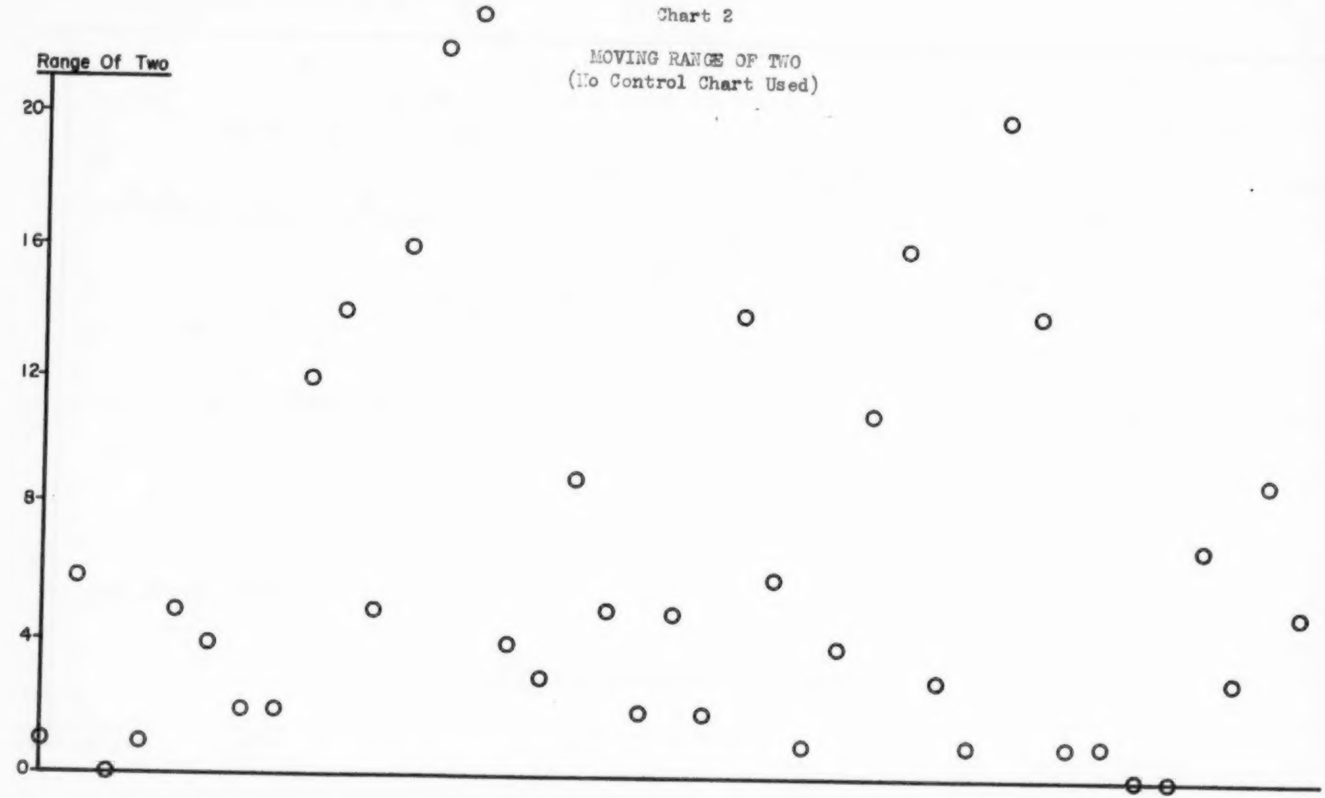


Chart 3

MOVING RANGE OF TWO

(Moving Average Control Chart Used)

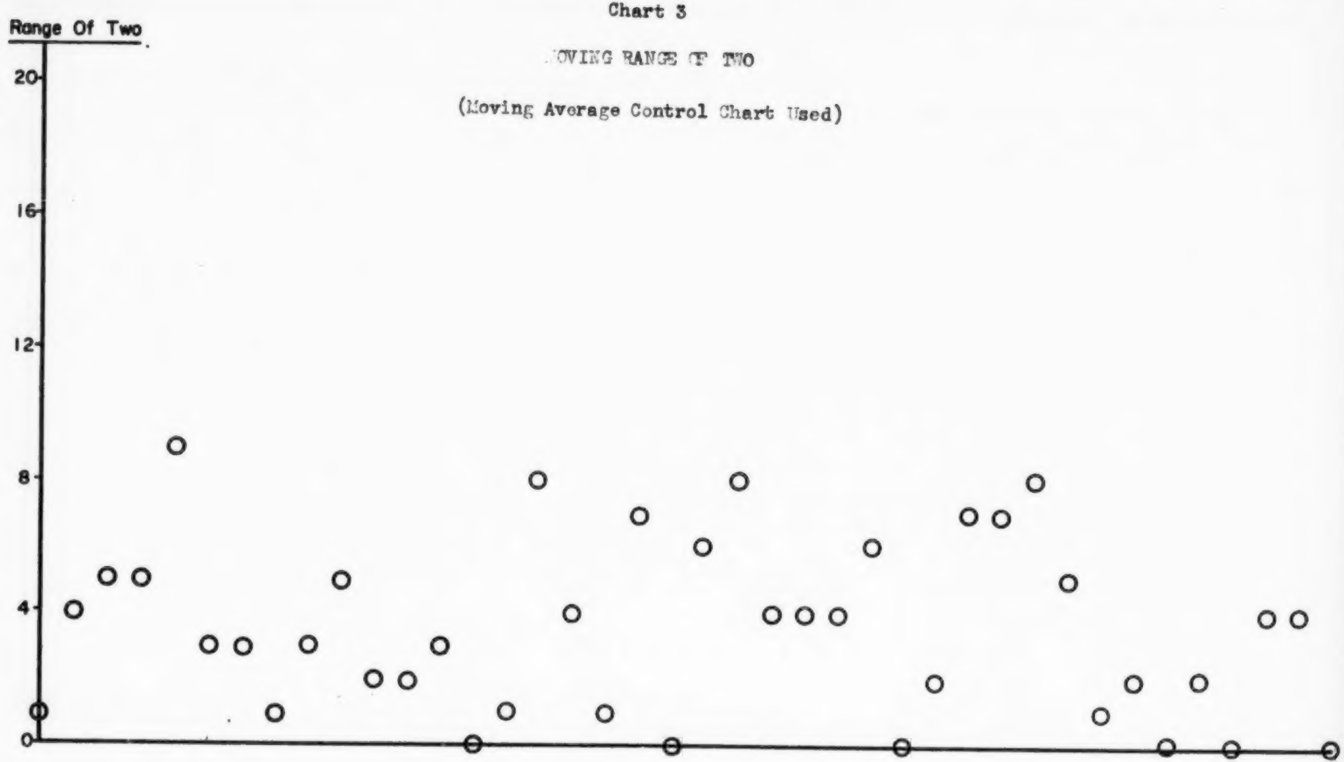


Chart 4

CONTROL CHART

(Moving Average of Four)

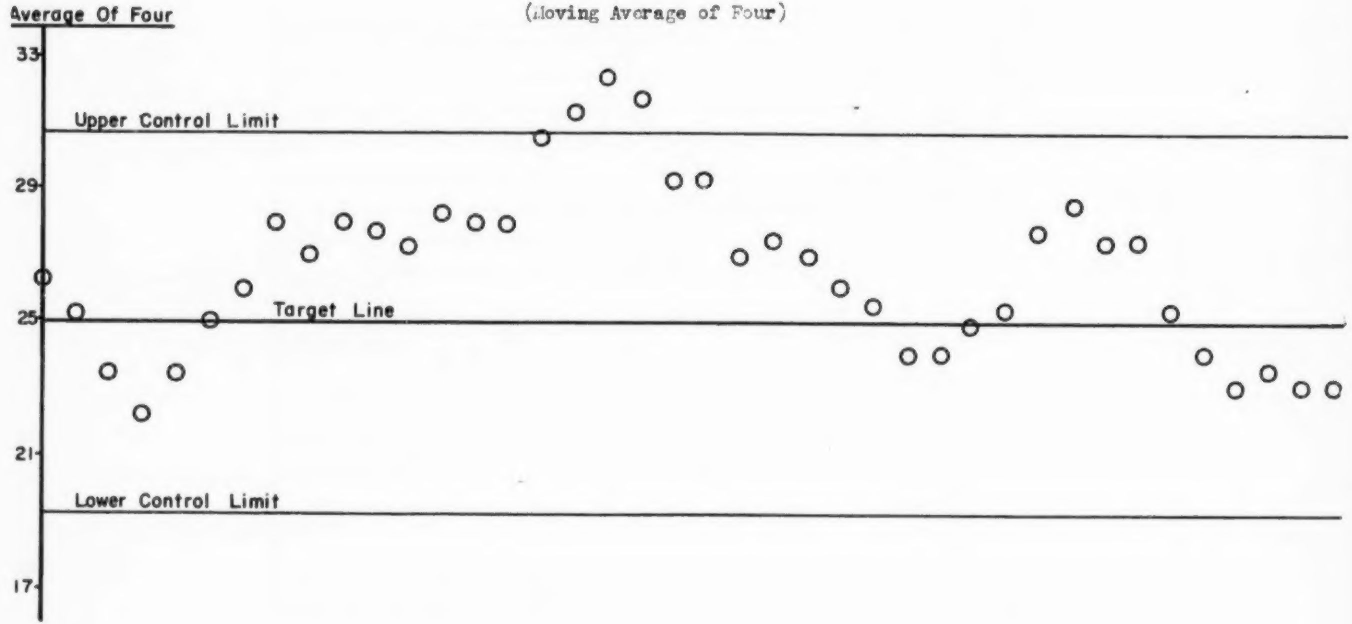


Chart 5

INDIVIDUAL SAMPLES

(Moving Average Control Chart Used)

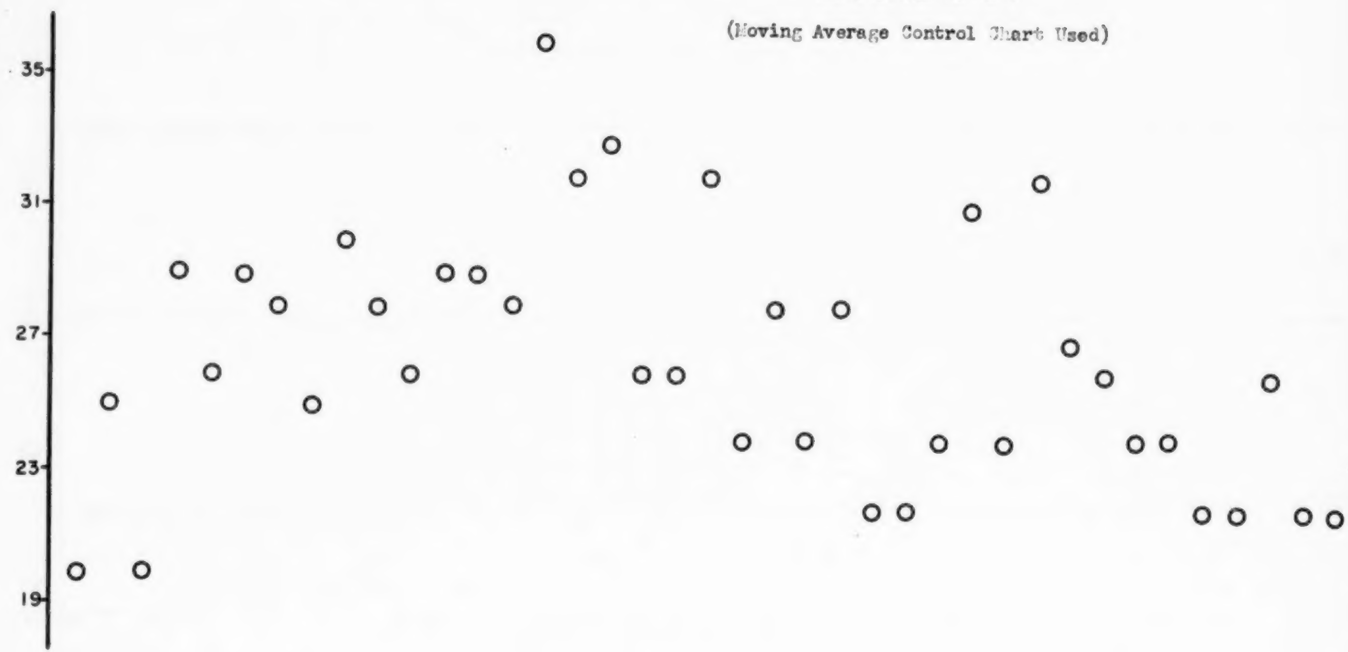


Chart 6

PROCESS ADJUSTMENTS PER FORTY SAMPLES

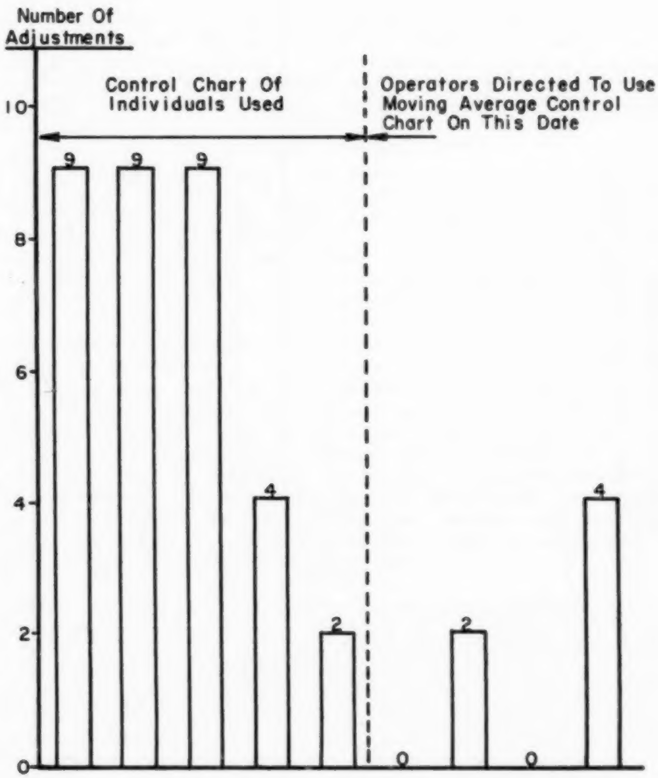
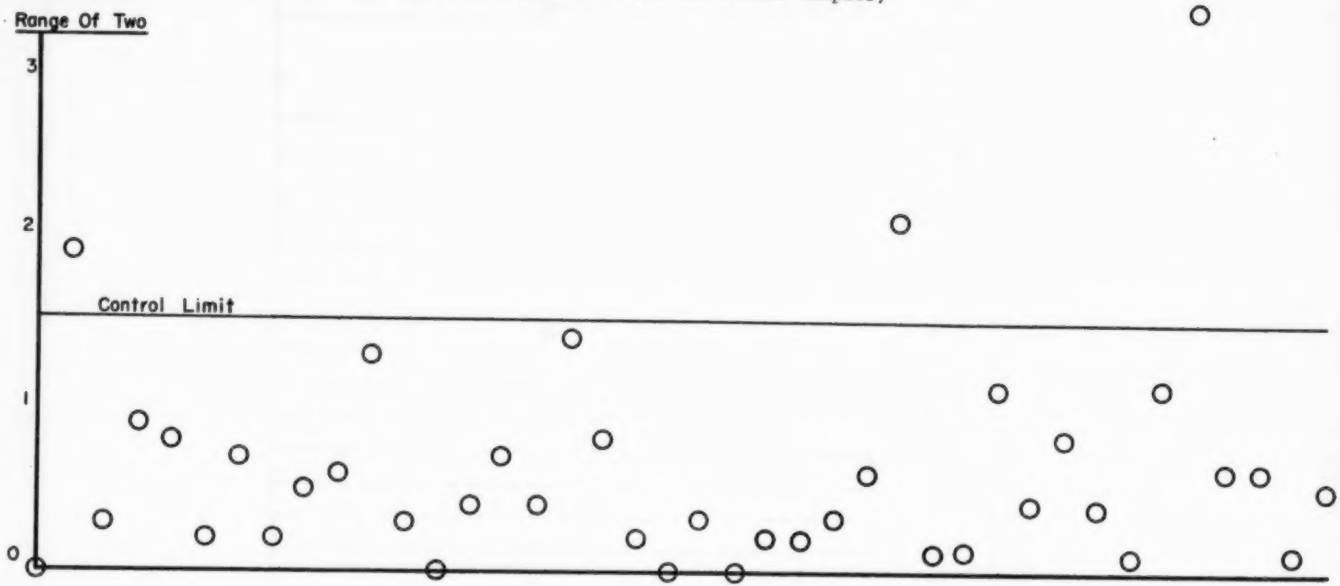


Chart 7

LABORATORY CONTROL CHART

(Range of Duplicate Tests on Routine Samples)



VARIABILITY PROBLEMS IN THE PHARMACEUTICAL INDUSTRY

Raymond L. Thatcher
E. R. Squibb & Sons
Division of Mathieson Chemical Corporation

The subject of variability is one with which we are all familiar. It is certainly a subject that receives considerable attention in the pharmaceutical industry.

In determining the potency of medicinals we become involved with analytical procedures that have many degrees of variability. Many of our chemical analytical procedures have insignificant standard deviations. On the other hand, some of our biological assays have large variances. For instance, in assaying insulin, we must inject over 5,000 mice to determine the potency, within desired confidence limits, of a single lot.

It is absolutely necessary to maintain variability of the physical characteristics of our products at an absolute minimum. Color variability is often improperly associated with potency variability. If tablets are not uniform in appearance, patients may feel, when they have prescriptions refilled, that the pharmacist has made a mistake. With these two examples it is not difficult to understand why we treat variability of physical quality characteristics with such respect.

We certainly have variability in our work. In our particular Quality Control laboratories, we have over 2,000 chemicals, intermediates and end-products that must be routinely analyzed and approved. In addition, container materials must be checked, packaging operations must be controlled, complaints must be evaluated, stability studies must be carefully scheduled, competitive products must be tested, raw materials from new sources must be compared, new analytical procedures must be checked or developed, specifications must be determined, legal requirements must be confirmed, appropriate shipping methods must be recommended and disposition of returned merchandize must be resolved. All related functions must be handled efficiently.

To carry out these varied functions and responsibilities it is necessary for our Quality Control Division to be staffed with a variety of personnel. To carry out our technical functions we have chemists, biochemists, pharmacists, chemical engineers, pharmacologists, pharmacognosists, bacteriologists, microbiologists, biologists, physical chemists, immunologists, statisticians. Many non-technical functions must supplement the technical. We have our clerks, typists, secretaries, samplers, porters, materials handlers, checkers, messengers.

It can thus be seen why we in pharmaceutical quality control cannot consider quality control synonymous with statistical quality control. To us, applied statistics is simply one of the tools, although a very powerful tool, that enables us to do a better job. Just as the infra-red spectrophotometer enables us to do better analytical work, so do statistical techniques better enable us to control our products and activities.

Therefore, although there are many non-statistical quality control problems that could be discussed, we shall limit ourselves to a discussion of variability from the viewpoint of applied statistics.

In the planning of our experiments and in the analysis of our data we need many of the variability techniques that have been devised. Control charts, chi-square analysis, analysis of variance, the F test, the t test, ranking methods and correlation techniques are all used. Although short-cut procedures are used to reduce work and speed answers, our calculating machines are daily pounding out sums of squares.

Applied statistics is, of course, not the exclusive property of quality control. Our Research Division is making extensive use of these procedures and has statisticians to supply the necessary service. Our manufacturing divisions are being serviced by our research and quality control statisticians. Our Production Control Division is receiving aid from us in operations analysis.

It is not the present intent to introduce any new mathematical theory or any new statistical technique. Two examples of how we use statistical techniques in solving some of our problems have been chosen. The first example has been selected because of its non-routine nature and the second because of possible interest in one of our routine biological assays.

We manufacture a drug which was assayed by a microbiological technique. The assay procedure was lengthy and time consuming. Two days were required before a potency value could be assigned to a lot. On each of two days we made ten individual determinations. We then averaged the twenty determinations to obtain the assigned potency. We knew that the standard deviation of these averages of twenty approximated 13.7 units. This had been determined by making 970 determinations on a single lot by performing ten determinations per day on 97 different days.

Our analytical research group investigated a quick chemical procedure. Fifteen determinations on a single lot indicated a standard deviation of approximately 7.3 units. It was, therefore, decided to apply the chemical assay procedure, in addition to the microbiological procedure, to each production batch for a trial period. Simultaneously, it also became desirable to evaluate the variability of the manufacturing process. The data obtained on sixty-two consecutive batches, which represented approximately one month's production, was evaluated in part in the following manner.

The microbiological and chemical assays (coded data) were first plotted (Chart A and Chart B). The microbiological results were distributed in a reasonably random manner with an average of 6.8 units and with the recorded upper and lower 95% probability limits. The chemical assay results were not distributed in a random manner since the number of runs below and above the median was lower than would be expected 2.5% of the time. The chemical assay data indicated a shift in level toward a higher average during the latter part of the period. In addition, the first ten lots assayed seemed to show remarkably low variability. The explanation for the non-randomness of the results, as is often the case when one is confronted with too few runs, is not completely explainable at this time. In spite of the non-randomness of the chemical assay results, it was decided, however, that the analysis of the data should be continued. Plotting of the data as histograms indicated reasonably normal distributions.

The next step taken was to determine whether correlation existed between the chemical and the microbiological assays. A correlation

factor of 0.06 was obtained indicating practically no correlation. This correlation value was found to be insignificant when checked by the formula

$$t = \frac{r\sqrt{N - 2}}{\sqrt{1 - r^2}} = 0.47.$$

Examination of the plotted data excluded the possibility of curvilinear correlation.

The variance of the microbiological results did not differ significantly from the variance of the chemical results.

$$F = \frac{(12.4)^2}{(11.7)^2} = 1.125$$

$$F\ (P = 0.95,\ \text{d.f.} = 61,61) = 1.53$$

The application of the t test to the difference of the two averages resulted in a non-significant difference.

$$P\left[-1.96 < z < 1.96\right] = 0.95$$

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = 1.84$$

The data was further analyzed in the following manner:

The variability of the chemical assay data was presumed to be the sum of the variability of the chemical assay plus the variability of the process from batch to batch. Similarly, the variability of the microbiological assay data was presumed to be the sum of the variability of the microbiological assay plus the variability of the process.

Thus

$$\sigma^2_{(C + P)} = \sigma^2_C + \sigma^2_P = 137.2$$

$$\sigma^2_{(M + P)} = \sigma^2_M + \sigma^2_P = 154.6$$

where

$\sigma^2_{(C + P)}$ = variance of chemical assay data

$\sigma^2_{(M + P)}$ = variance of microbiological assay data

σ^2_P = variance of process

σ^2_M = variance of microbiological assay

σ^2_C = variance of chemical assay

The chemical assay on each of the sixty-two lots was subtracted from the corresponding microbiological assay. The variance of the sixty-two differences was calculated to be 291.0.

$$\sigma^2_{(M + P) - (C + P)} = \sigma^2_{(M - C)} = \sigma^2_M + \sigma^2_C = 291.0$$

By calculation

$$\sigma^2_{(M + P)} - \sigma^2_{(C + P)} = \sigma^2_M + \sigma^2_P - \sigma^2_C - \sigma^2_P = \sigma^2_M - \sigma^2_C$$

$$= 154.6 - 137.2 = 17.4$$

Then solving for σ^2_M and σ^2_C by using the simultaneous equations

$$\sigma^2_M - \sigma^2_C = 17.4$$

$$\sigma^2_M + \sigma^2_C = 291.0$$

We obtain

$$2\sigma^2_M = 308.4 \qquad 2\sigma^2_C = 273.6$$

$$\sigma^2_M = 154.2 \qquad \sigma^2_C = 136.8$$

$$\sigma_M = 12.4 \qquad \sigma_C = 11.7$$

and

$$\sigma^2_P = 154.6 - 154.2 = 0.4$$

$$\sigma_P = 0.63$$

It was thus found that the variability of the chemical assay data and the variability of the microbiological assay data were due almost exclusively to the variability of the assay methods. The process itself was exceptionally uniform. This also explained why practically no correlation was found between the two assay methods.

The result of this analysis was the discontinuance of the microbiological assay on a routine basis. Although the expected variability of the chemical assay was not realized, it was decided that under routine conditions it was as equally reliable as the microbiological procedure and that additional experience in the hands of our analysts would ultimately result in reduced variability.

In the assay of corticotrophin, commonly known as ACTH, as with a number of pharmaceuticals, it is not possible to measure directly the pharmacological effectiveness. Neither can the product be assayed chemically, nor the amount of the effective factor determined. A secondary biological effect is, therefore, measured.

In this case it has been the practice to determine the depletion of ascorbic acid in the adrenal glands of hypophysectomized rats. This has been shown to be related fairly linearly to the log dose of ACTH administered. To allow for the normal expected variation in the size of the adrenal, even in carefully selected rats of nearly identical age, weight, and heredity, the concentration of ascorbic acid per unit weight of gland has been the characteristic used rather than the absolute amount of ascorbic acid. A typical set of response curves is shown in Chart C. The values obtained on individual rats are indicated and show the wide variation experienced, even with the use of carefully selected animals.

The potency of a lot is determined by comparison of the log dose response line to that of a standard material by usual factorial methods. Even here, however, modern short-cut methods have come to our aid. Formulae have been developed for the use of the range of responses at each dose level as an estimate of variation so that it is not necessary to go through the longer computation of the standard deviation.

We use V as the difference between the sum of all the responses of the unknown, and the sum of all responses of the standard; and W as the difference between the sum of responses to the high dose of both materials and the sum of the responses to the low dose. The usual equation as outlined by Bliss, Emmens, etc., for the potency of the unknown is expressed in terms of the per cent of the assumed potency as

$$\text{Potency } (\%) = \text{antilog } \left(2 + \frac{4\ i\ V}{W}\right) = A$$

and the standard error of the assay is written as

$$S = \frac{4.340\ A\ i\ \bar{R}\sqrt{n}}{D\ W^2}\sqrt{3\ W^2 + 2\ V^2}$$

where i = log dose interval

$\bar{R}$ = average range

n = number of animals on each dose

$D = d_2$, the usual factor for estimating standard deviation from $\bar{R}$.

Since a usual value of the standard error for such an assay, using eight rats on each dose of both the standard and the unknown, is in the neighborhood of 15%, it is necessary to run several assays to obtain an estimate of the potency within desired confidence limits. The final potency of a lot is obtained by averaging the individual assay potencies, each weighted by the reciprocal of its standard error. The combined standard error is computed by the formula

$$S_c = \frac{1}{\sqrt{\frac{1}{S_1^2} + \frac{1}{S_2^2} + \text{---} \frac{1}{S_n^2}}}$$

where S_1, S_2, --- S_n are the standard errors of n repetitions of the assay.

There are days when it is evident that some unassignable cause changes the characteristic slope of even the standard so as to render the assay valueless. Whether this can be laid to special climatic conditions, to different life histories of the animals or what, is not known, but it does occur. To determine when assay results are sufficiently irregular to warrant scrapping the assay, our laboratory employs a modified control chart for the slope of the log dose response line. Based on many months' data, three sigma control limits have been established. When the slope of a response curve falls outside these limits it is considered that some unassignable factor is present and the assay is discarded. A portion of the control chart is shown in Chart D.

In the use of this assay, various dose levels and log dose intervals have been tried. In our laboratory we have considerably lowered the general dose level as the result of the use of this control chart and associated analyses. This was done because it was discovered that the slope was greater at lower doses and that the standard deviation of response about the average for the lower dose was no greater. This results in a more satisfactory ratio of the slope to the standard deviation about the average at any point. The precision of the assay is thereby increased.

The use of the smaller doses has resulted in further investigation. As has been mentioned the concentration of ascorbic acid in the adrenal gland has been used. Laboratories abroad have found some evidence, and it has been suspected in our laboratory as well as in others here, that what we are measuring is not a single effect but a combination of effects, a fact that is only of importance at low dose levels. It is now considered possible that ACTH not only acts to deplete the ascorbic acid in the adrenal gland, but also tends to decrease the weight of the gland itself. At high dose levels these effects are probably parallel but at low dose levels they do not appear to be, so that investigations are being carried out in an attempt to determine whether, if the animals are carefully selected, the absolute value may not be the better factor to use. This is a matter which must be considered statistically, to determine whether the response lines are linear and of sufficient slope, and whether variability about the average response at a dose level is no greater for the absolute value than for the concentration.

In addition, comparison of the standard errors of the final potencies based on the two methods must be made. Inasmuch as using the absolute value will shorten and simplify the assay procedure, it will be advantageous to show that its use is equally satisfactory.

Under conditions of great variability as experienced in this and other biological assays, it would be difficult to reach a decision without the use of modern statistical techniques and an understanding of probability distributions. Wide variability makes it particularly difficult to assess the value of small differences in average values unless such resources are relied upon.

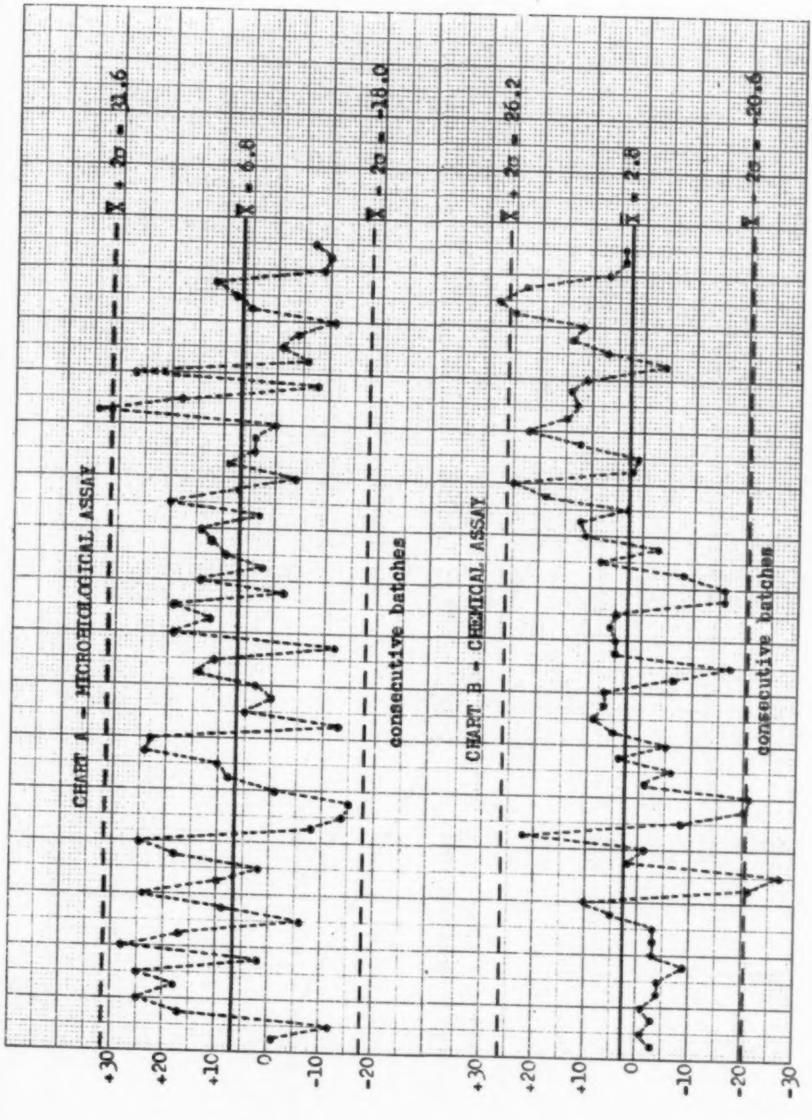


CHART A - MICROBIOLOGICAL ASSAY
$\bar{X} + 2\sigma = 31.6$
$\bar{X} = 6.8$
$\bar{X} - 2\sigma = -18.0$
consecutive batches
CHART B - CHEMICAL ASSAY
$\bar{X} + 2\sigma = 26.2$
$\bar{X} = 2.8$
$\bar{X} - 2\sigma = -20.6$
consecutive batches
+30
+20
+10
0
-10
-20
-30

CHART C - ACTH ASSAY

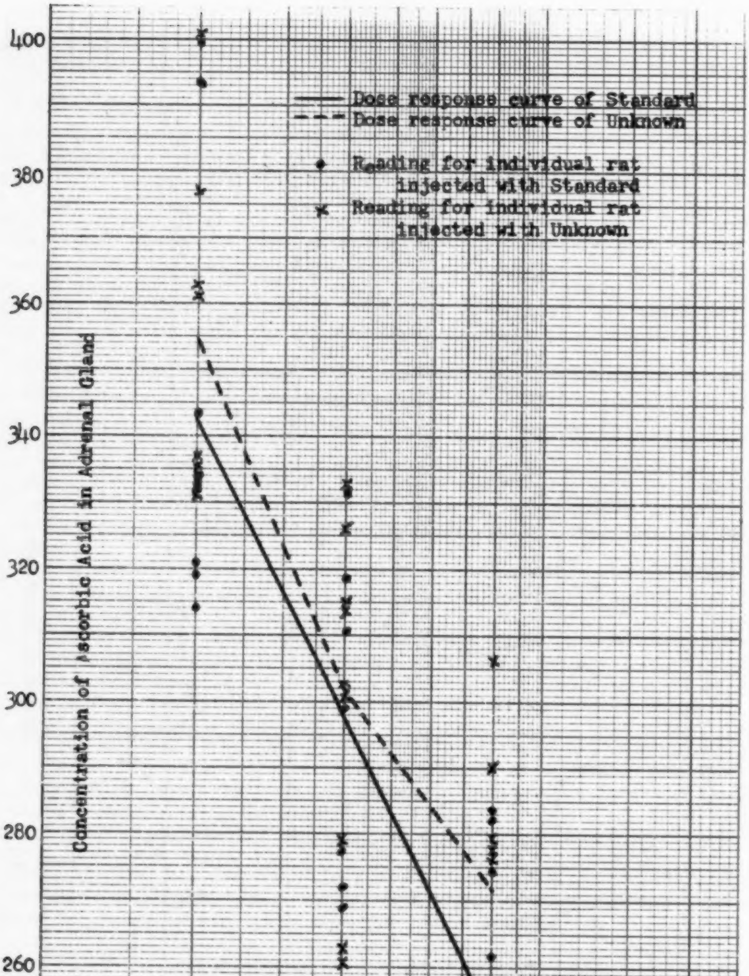
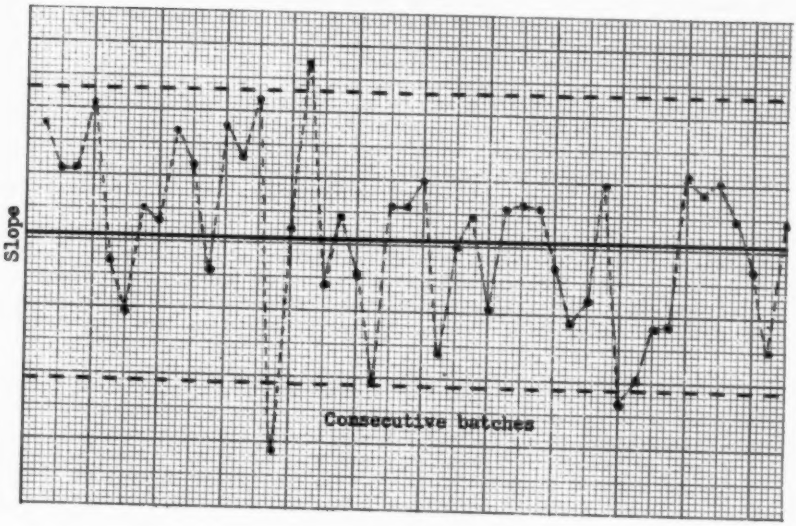


CHART D - ACTH ASSAY

Control Chart for Slope of Response Curve
Two σ Limit Lines



References:

"Simplified Range Method for Quantitative Responses such as ACTH" - U.S.P. - ACTH Memorandum 5 by Lila Knudsen.

"Principles of Biological Assay" by C. W. Emmens.

"Estimating Precision of Measuring Instruments and Product Variability" by Frank E. Grubbs, Journal of American Statistical Association, Volume 43, 1948.

IMPROVING THE QUALITY OF INCOMING MATERIAL

Stephen J. Rogers
Ford Division of Ford Motor Company

It is indeed a pleasure to be here today to discuss with you the Receiving Inspection and related activities of the Ford Division of the Ford Motor Company. We will all agree that, in order to guarantee and sustain the quality of his product, every manufacturer must first be certain his basic materials are of top quality. It is doubly essential to assemblers of material like ourselves, since everything but skill in assembly depends on material received from outside sources, and sometimes I think even skill in assembly depends on the quality of parts and material we provide the man on the line. To some degree, each of us is involved in the problem of determining the quality of incoming material and in the more complex problem of improving it. We hope our approach to these problems we share will prove interesting to you.

Several years ago, we took a long critical look at our methods in Receiving Inspection. We recognized two basic functions of this operation, namely, to prevent the use of defective material and to produce facts necessary to effect prompt correction of material at the source. We determined that we could improve our execution of these functions and our use of the facts by:

1. Establishing a firm schedule of inspection for each part. The number of parts used in a car plus low inventories requiring frequent repetitive shipments make it impractical economically to inspect every shipment. Furthermore, the increasing acceptance of statistical control methods demonstrates the advantages of a controlled sampling plan.

2. Assigning the selection and classification of inspection characteristics to qualified process engineers specializing in particular types of material, parts, or assemblies.

3. Recording inspection results in a flexible form which would lend itself to review, comparison and discussion with a minimum of delay.

4. Originating a common fixed language for use between ourselves and the supplier to promote ready understanding and proper interpretation of inspection rejection reports and findings.

5. Obtaining personnel qualified to analyze a manufacturing process, determine its faults and make recommendations to the supplier. We recognized that opportunities to assist our suppliers in producing quality products existed subsequent to the time when we furnished them with a blueprint and material specifications or with a rejection report on one or more of their shipments. We wanted to implement the statement that the best place to guarantee quality is in the supplier's plant.

Our first step was to hire a group of men we called Material Technical Specialists. Each man is an expert in a particular field, such as chemical engineering, engine manufacture, gear manufacture, textile

manufacture, etc. We established eight categories of material, assigning responsibility for each category to a Material Technical Specialist.

Next we designed an Inspection Operation Sheet on which to detail the characteristics to be inspected, the method of inspection and the schedule of inspection. At this point I'd like to show you our first slide which gives the initial step in development of an Inspection Operation Sheet.

When the part prints are first released by Engineering, the Technical Specialists process them, selecting the points to be inspected, marking and classifying them by order of importance as safety, critical, major, minor and incidental characteristics. The highest order of individual characteristic within the part determines the class of the part itself. For example, a part containing major, minor and incidental characteristics is called a major part.

Definitions of characteristic class are as follows:

Safety - failure of these characteristics may endanger the individual using or maintaining the product or cause the user to jeopardize the safety of other parties.

Critical - failure of these characteristics may prevent or seriously impair the functioning of the product or may result in high cost to replace or repair after assembly.

Major - failure of these characteristics may materially reduce the usability or life of the product, or affect adversely the appearance of the product, or may result in serious assembly or repair problems.

Minor - failure of these characteristics may result in assembly and repair problems or affect the appearance of the product.

Incidental - failure of these characteristics may result in minor problems but will not seriously affect the function of the end product.

This slide shows an Inspection Operation Sheet written from the marked print by Quality Control Analysts who group the characteristics by classification, describe them, assign the frequency of inspection to each characteristic class and determine the frequency of laboratory and layout inspection. Receiving Inspection of safety and critical characteristics is scheduled for every shipment. Receiving Inspection of major, minor, and incidental characteristics is scheduled in accordance with the nature of the characteristics and the shipping schedule of the supplier. For example, major, minor, and incidental may all be placed on a frequency of one if only one shipment a month is scheduled from the supplier. The frequency assigned to Laboratory and Layout Inspection depends on the nature of the part and on the shipping schedule of the supplier.

As shown by the Operation Sheet, the minimum number of inspections given any group of characteristics of this part in 30 shipments, is three.

The first shipment is checked completely, each shipment thereafter is controlled by the frequencies of the Inspection Operation Sheet. The results in number of inspections across 30 shipments is as follows:

Receiving Inspection

Safety & Critical characteristics	-	checked	30	times
Major	" -	"	7	times
Minor	" -	"	4	times
Incidental	" -	"	3	times
Laboratory Inspection	-	"	4	times
Layout Inspection	-	"	3	times

Two other factors bear on the amount of inspection given each part. First, if a shipment is rejected, the schedule immediately reverts to a frequency of one for the particular group of characteristics or type of inspection involved and remains at one until released by an accepted shipment. As an additional precaution, if a characteristic not scheduled for inspection is found defective in assembly, a special inspection is performed on the shipment involved. If rejected, the group of characteristics involved is placed on a frequency of one until released by an accepted shipment.

In addition to establishing the "When" and "What" of Inspection, the Inspection Operation Sheet covers the "How." When the Operation Sheet has been written up to this point, it is reviewed by a Gage Specialist who forwards his recommendations to our Gage Design and Procurement Unit. When gages have been agreed on, designed and procured, the gage identification number is added to the Operation Sheet.

Some parts do not lend themselves to the normal type of Inspection Operation Sheet because of complexities arising from repetition of characteristics in different locations. Our next slide illustrates a special purpose Inspection Operation Sheet for the Tudor Side Panel. Here we have pictured and coded weld patterns, sub-assembly fits, and the locations and shape of important holes. In addition, we have coded appearance defects such as dents, dings, die marks, buckles, etc., and provided a coded location pattern. With the aid of charts like these, we are able to standardize inspection and reporting at widely separated points. Most important of all, we establish an easily interpreted means of communicating with our suppliers. The next slide shows a typical report of appearance defects. This report is sent to the Quality Control Manager of our supplier.

We have applied this method to large metal stampings, such as hoods, roofs, and side panels, and to glass and seat springs. The next two slides show our charts on windshield glass and seat springs.

Gentlemen, now that we've covered the Inspection Operation Sheet which serves as our primary tool in inspection, we will set forth our operations from the time a part is officially released for procurement through its routine inspection during production. The 1954 Ford contains over 1800 new parts with approximately 1.4 suppliers per part. Prior to production, each new part from each supplier must be checked to material specifications and the blueprint. Our next slide shows the flow of the Initial Sample procedure. The Purchasing Department authorizes a prospective supplier to make and ship a sample, usually six pieces, from

initial tooling. The sample is subjected to a Laboratory and Layout check for conformance to material specifications and configuration. Purchasing is notified of approval or rejection. If approved, the supplier is instructed to harden his tools and prepare for production. If rejected, the supplier is given the reason and asked to effect correction and submit new samples. In some instances as many as three or four samples are submitted before final approval.

Our next step is to receive a sample from the production tools or processes. This is called the First Production Shipment. Our next slide shows the flow of this procedure. Production Control authorizes shipment from the suppliers first production run. Samples from this shipment are inspected to the Operation Sheet by Receiving Inspection, to the blueprint by Layout, and to the material specifications by the Laboratory. In addition, a functional test is made to determine how the part will fit with its mating parts and perform on the vehicle. Production Control is notified of approval or rejection. If approved, the supplier is instructed to begin production in accordance with the Production Control Schedule. If rejected, the supplier is given the reason, asked to effect correction and submit new production samples. In some instances, several attempts are made before the part is finally approved.

After approval of the First Production Shipment, we enter the routine program of sampling the part during its production life. Our next slide illustrates the application of schedules for inspection which were established on the Inspection Operation Sheet. For each part we have an IBM card containing part name and number, supplier name, supplier shipping point, buyer code, purchase order number, material classification and an inspection schedule for Receiving Inspection, Laboratory, and Layout. The card is designed for tallying each shipment received in the blocks at the right. From the progressive tally and the inspection schedule, the Cardatype Operator is able to determine what inspections to schedule.

This is a packing slip which accompanies each shipment. After the shipment is checked in by the receiving people of the plant, the packing slip is forwarded to the Cardatype Operator. He pulls from his file of cards the Master Card applicable to this particular supplier, shipping point, and purchase order number. He next tallies the shipment and determines what types of inspection are required. In our illustration, the tally number is thirty, the schedule is one, or every shipment, for Receiving Inspection, every 10th shipment for Laboratory, and every 15th shipment for Layout. Therefore, the operator must schedule all three types of inspection. The card is inserted in the slot shown here on the right hand side of the Cardatype machine. The machine automatically types the supplier name, buyer code, part name, part number, shipping point, purchase order number, supplier code, material classification, and date received. The operator manually types the variable information, quantity received, inspection report number, packing slip number, inspection code, date shipped, method of shipment, car number and waybill number. This operation is repeated for each type of inspection. While the forms are being typed, a tape is automatically punched recording the part number, supplier code, material class, inspection report number, date received, quantity received, and type of inspection. The tape is used in another machine called the Tape-to-Card Punch which automatically produces the inspection follow-up cards shown at the bottom of the slide. These cards are used to follow-up on each scheduled inspection and to provide a daily record of the number and types of inspection scheduled in each material class.

In Receiving Inspection we employ a double sampling system which encompasses sampling by attributes. JAN Standard 105 at inspection level 3 is used as a basis for the sample size and acceptance number. The Inspection Operation Sheet calls out the characteristics for each inspection, and is keyed to the Inspector's Report by the Inspection Code which states the consecutive number of the shipment. As our methods of actual inspection are common to most businesses, we will not go into detail about them.

We mentioned earlier that we are most anxious to actively assist our suppliers in correcting their product at the source rather than simply preventing defective material from being used in production and reporting its existence to the supplier. It was stated that we hired men who are expert in specific fields. I want to emphasize their qualifications. They are capable of setting up a manufacturing process or analyzing it in detail.

We could not simply hire a Technical Specialist, however, no matter how well qualified, and expect him to perform his job without tools. It is the need for those tools which has stimulated much of the program we have discussed with you. The precise definition of inspection contained in the Inspection Operation Sheet, the standardization of nomenclature also inherent in the Operation Sheet, the localization of defects made possible by the special charts on stampings, etc., are tools designed to bring <u>to the Specialist</u> the facts concerning defective material. Armed with the facts, he is in a position to visit the supplier's plant and discuss corrective action programs.

Our next slide deals with the flow of information to the Technical Specialist. Copies of completed Inspection Reports are forwarded to the Tabulating Unit and the results of inspection are entered on the Inspection Follow-Up Card. Each day cards representing rejections are listed by material classification. The lists which serve as a register for following-up and recording action, are forwarded to the Specialist accompanied by the Inspector's Reports. It is then the responsibility of the Specialist to take action on each rejection. First he must contact the supplier, inform him of the rejection, the cause, and determine what corrective action the supplier intends to take. If the supplier is aware of the problem, he may have a corrective program established. For example, if a punch is broken and has caused ragged holes, it is possible that it has been discovered and sharpened. In this case, the supplier would know when good production started from that operation and could give us an effective date. If he is unable to give a corrective date because he is not aware that his material is of poor quality, he is requested to investigate the problem and inform us of the action he is going to take and give us the date that action will be effective. In many cases a conference at the supplier's plant or at Ford is required to discuss the problem and methods of correction.

Next, the Specialist must inform the affected assembly plants that shipments in the pipeline from a particular supplier must be inspected for the specific defect involved. In addition, a Material Quality Status Report is prepared. Publication of this report provides to each assembly plant the status of parts with which they might have difficulty and permits them to refer to the status of parts which have given trouble in the past. The Material Quality Status Report is cumulative for each characteristic of a part, as illustrated.

Our next slide pictures two monthly reports which we use extensively in discussions with our suppliers. The first, the "Part Quality Record," is a detailed, chronological report of all shipments received during the month, all inspection and all action taken by our Specialists. Defects found in Receiving Inspection are expressed in codes derived from the Inspection Operation Sheet. Laboratory Inspections are identified in the same field with the number "22". Layout Inspections are identified with the number "33". Inspection disposition is given in the last column. 1 indicates approval; 2, Rejected-Usable; and 3, Rejected-Not Usable. Receiving Inspection for the month is summarized for each part number-supplier and given as the number of pieces received, number of pieces inspected, number of defective pieces, number of shipments received, number of shipments inspected, and number of shipments rejected.

The second report, the "Supplier Quality Record," lists for each supplier all parts supplied and gives a three months' summary of his performance. This report gives us a concept of the quality the supplier has been producing over an extended period. We will have more to say about the major use of this information in a moment.

Many of you no doubt are familiar with the difficulties involved in obtaining complete correction of sub-standard material. There is a multitude of factors which can and do intrude between the simple statement by a consumer that material received from a supplier is not to specifications and an equally simple action on the part of the supplier to correct the fault. These factors run the gamut from raw material difficulties to the complexities of modern machine tools.

I wonder how many of you have been given definite assurance that an offending defect will be corrected, only to receive the next shipment after the alleged corrective date in the same condition? When this happens, we may follow any one of several courses of action. We may reject every shipment, providing our production is not affected. Or we may sort every shipment at the supplier's expense if production is affected. We may change suppliers. Yes, we have courses of action open and we use them when necessary. But this should not be the sum total of our program or we are committed to operating on an emergency basis, rushing from fire to fire with our extinguishers. The supplier must be made aware of the problems inherent in the parts he makes and the importance of actively assisting us by maintaining effective control over the characteristics that are important, based on experience.

Let's review briefly our first attack on this problem. A Technical Specialist, fully qualified, contacts the supplier when material received is determined to be sub-standard. He is in possession of the facts. If the defect is of unknown origin or is technically difficult to correct, he is available to suppliers for consultation; he will work with the supplier in his plant to determine the cause and assist in the correction. Normally, you would expect that this would be sufficient. Actually, in many cases it is not.

Realizing then, that production and quality control personnel of a supplier are sometimes unable to act decisively, we have added one more step to our procedure. Using the "Supplier Quality Record" and "Part Quality Record" we review the performance of each supplier each month. If the record indicates that the supplier has submitted defective material on one or more parts, has been contacted and has failed to effect correction, or has had intermittent success in controlling a troublesome

characteristic, we request Purchasing to arrange a meeting with the supplier's management to discuss the problems involved. These meetings are attended by our Specialists and by representatives of our Purchasing Department.

Our discussions are not limited to the defects we found in his last shipment. We may discuss his equipment, inspection methods, raw material or any factors which bear on his performance. We describe our methods and our objectives. Above all, we try to enlist his full and active cooperation in our endeavor to guarantee quality in his plant, to our mutual benefit.

Our experience has been that once fully aware of the extent and seriousness of a problem, management of our suppliers move swiftly to control it. We feel that the results of our meetings will be long-lasting. We have many indications that they will put their newly acquired knowledge of our operations to good use; that they will want our Inspection Operation Sheet for revising the emphasis of their own inspection; that when machine or process changes are necessary they'll ask for our prior approval of a production sample; and that they will take greater advantage of the technical advice we are able to offer. I wish to emphasize one point in this connection and that is that our system of quality control is designed to assist suppliers in producing high-quality products rather than to exonerate them for defective and unsatisfactory work. In other words, the ultimate responsibility for quality in all cases rests with the supplier. We do not undertake to share this responsibility merely by instituting a system of inspection at the source; rather we intend only to assist each supplier in doing a better job in manufacturing products for us. Purchase orders which we issue contain a provision in accordance with which suppliers make certain guarantees with respect to the quality and suitability of products sold to us. The inspection processes which we carry out are not designed to be a substitute for such guarantees or to operate as a waiver of our rights in connection with them.

Our last slide gives tangible evidence of the effectiveness of our approach. In a period of increasing production and correspondingly increasing numbers of shipments as indicated by the first line on the chart, our percentage of rejections dropped substantially as indicated by the second line.

This concludes our presentation. In it we have sought to show how we have departed from the usual concept of Receiving Inspection to a much broader organization calculated to promote cooperation, coordination and understanding between consumer and supplier. Much remains to be done and we are constantly revising and improving as we gain experience. We firmly believe that an orderly, controlled plan of inspection plus the technical know-how to assist a supplier in solving his quality problems can be very effective. We feel that this combination can contribute the most to our common objective - the prevention of the manufacture or release of sub-standard material at the source.

STATISTICAL METHODS APPLIED TO STEEL PLANT OPERATIONS

Arthur P. Woods, Jr.
Armco Steel Corporation

Among the many important groups of problems facing the management of a steel plant, there are two whose solution can be more closely approached by studies involving the use of statistical methods. These problems are:

1. How to use the presently available plant equipment most efficiently.

2. How to spend any money available for plant improvement and expansion most effectively.

The studies necessary to solve these problems often require laboratory, pilot plant, or production plant experimentation. The use of statistical methods increases both the amount and accuracy of the information obtainable from such experimentation.

In any manufacturing process raw materials (e.g., iron ore, coal, steel scrap), after entering the plant, undergo a series of changes in various pieces of equipment (e.g., blast furnace, coke ovens, open hearth furnace, rolling mills) and finally leave the plant as a finished product (e.g., steel coils and sheets). At each step in the operation the condition of the equipment being used and the product being processed may be defined by a number of operating and product factors. These are of two kinds:

1. Variables are factors whose location is known and whose magnitude is measured (e.g., rate of open hearth furnace fuel consumption in BTU/hr.).

2. Attributes are factors whose location is known but whose magnitude is not measured.

In an investigation of a process attributes are defined by stating their location and studied by comparing them with other attributes serving the same purpose (e.g., comparing three different designs of open hearth fuel burners, comparing two different sources of the same raw material).

The operating factors (those which define the equipment and its operation) differ from one another in the ease with which they may be controlled at a desired level by the operator.

1. Some factors are under the control of the operator within the limits of the equipment or the supply available (e.g., the rate of fuel consumption).

2. Some factors can be made more controllable after the addition of more capital equipment (e.g., the rate of scrap charging into the open hearth furnace).

3. Some factors are controllable only indirectly by controlling other operating factors (e.g., the preheat temperature of the combustion air entering the open hearth furnace chamber).

All product factors (those which define the condition of the product in or flowing through the equipment) are controlled indirectly (e.g., the chemical analysis of the steel produced). The assumption is made that variations in the product factors are due to variations in the operating factors and in the raw materials entering the plant.

The efficient use of the presently available plant equipment implies that the product factors are being controlled at a desired level that gives a maximum profit to the company while supplying satisfactory quality to the customer. A study of presently existing operations to learn what product and raw material factors cause variations in production rate, quality and cost, and how these factors may be controlled gives management much of the information it needs to operate the plant efficiently.

The effective spending of plant improvement and expansion funds implies that customer needs for more product, better quality, lower prices or new products are being met in a way that will give maximum company earnings for each equipment dollar spent. The following types of investigations will furnish some of the information management needs if effective spending of funds is to occur.

1. A study of presently existing operations to learn what equipment bottlenecks exist that would prevent improvement in production rate, quality or cost reductions should changes in operations be made.

2. A study of the fundamentals of each piece of equipment to learn its most efficient design (e.g., model studies of open hearth furnaces involving dimensional analysis).

3. Trials of new processes or pieces of equipment to compare their performance with present operations (e.g., the development of the continuous hot dip process for making galvanized steel coils to replace the old hot dip sheet process).

4. Development of new products and processes to make them (e.g., the development of grain oriented silicon steel for transformer cores).

5. A study of the chemical, physical or mechanical properties of products or processes to obtain a better description of the product or a law governing the process. This information is then used in investigations of type 2, 3 and 4 mentioned above (e.g., the physical properties of steel, new analytical methods, and thermodynamic properties of elements, compounds and reactions).

There are two types of experimental procedures used in these investigations.

1. The controlled experiment where all or most of the factors being studied are held close to pre-arranged values. The effects of the various factors on the dependent variable being studied can be easily obtained by analysis of variance methods (1, 2).

2. The uncontrolled experiment where most of the factors being studied vary over a range of values. The data obtained must be analyzed by correlation methods. Simple or multiple correlation will show the average effects of the independent factors, but the correlation methods must be modified to show interaction.

Interaction between two independent variables exists when the effect of one of them on quality, for example, changes when the level of the second independent variable is changed (e.g., the effect of sulfur on the hot rollability of steel changes as the percent manganese present decreases).

The following example of an uncontrolled experiment on a presently existing operation will illustrate how correlation methods may be modified to show important interaction effects.

Data for this example were selected from a study made to determine the effect of certain operating and product factors on basic open hearth production rate, measured either in tons per hour or heat time. The original study contained data on 38 factors collected on 130 heats made in a single furnace campaign at the Middletown Division's No. 1 shop.

The raw materials charged into a furnace undergo two processes during the making of a heat of steel.

1. The scrap is melted and the steel heated to the proper tapping temperature.

2. The carbon, manganese, silicon, sulfur and phosphorus added with the raw materials are oxidized out of the molten steel until the desired analysis has been reached. Adjustments of analysis are then made by adding alloys to the steel either in the furnace just before tap or to the ladle.

A comparison of the relative stage of these two processes was obtained in the experiment by taking an early carbon test (called the melt carbon test) from the furnace during the period between one hour after flush to start of lime boil.

In this example, the following factors were studied to determine their effect on production rate:

1. Percent hot metal.

2. Percent melt carbon.

3. Scrap charging rate.

The following curves, Figure 1, show the results obtained when graphical multiple correlation is used. The independent effect of each factor is shown, but any interaction between the factors is averaged out. In each of these curves the carbon at tap has been graphically corrected to 0.06 percent.

The graphical multiple correlation method, as applied to individual samples, is thoroughly explained in Chapter 16 of the reference (3). The method becomes rather tedious when applied to 130 samples. However, the data may be grouped to reduce the amount of work required.

In preparing the hot metal curve shown in Figure 1, for example, the 130 individual samples were sorted in order of increasing Percent Hot Metal and divided into 13 groups of 10 heats each. The average Tons per Hour, Percent Hot Metal, Percent Melt Carbon and Scrap Charging Rate was obtained for each of the 13 groups. The 13 Tons per Hour values were

FIG I. THE EFFECT OF PERCENT HOT METAL, PERCENT MELT CARBON AND SCRAP CHARGING RATE ON OPEN HEARTH PRODUCTION RATE IN TON PER HOUR

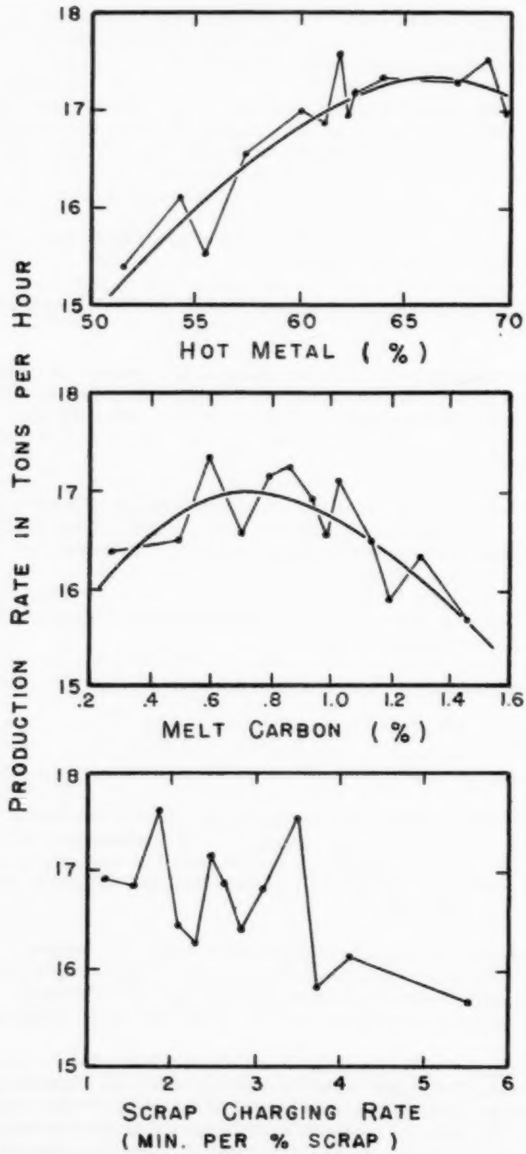


FIG 2. THE EFFECT OF PERCENT HOT METAL ON THE RELATIONSHIP BETWEEN PERCENT MELT CARBON AND OPEN HEARTH PRODUCTION RATE IN TONS PER HOUR

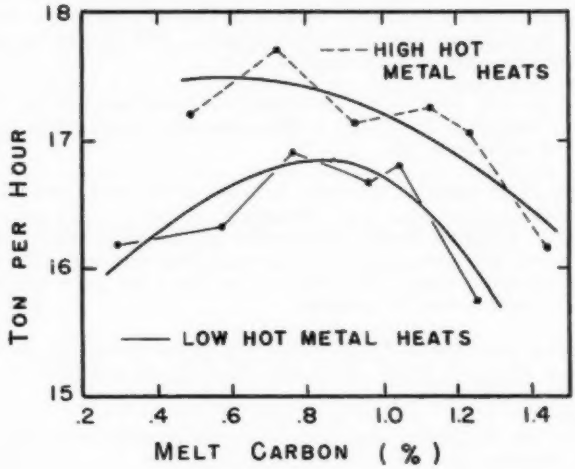
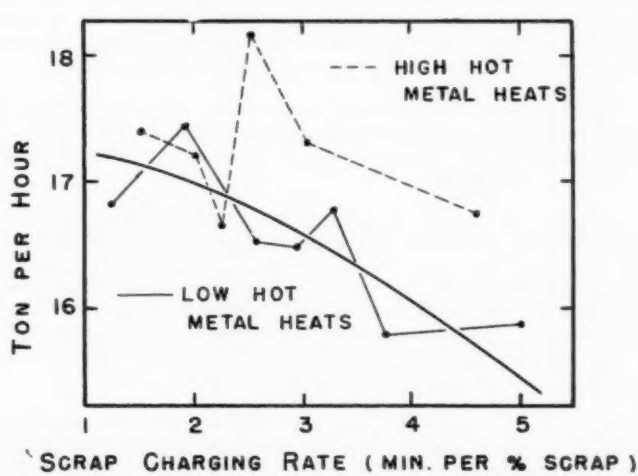


FIG 3. THE EFFECT OF PERCENT HOT METAL ON THE RELATIONSHIP BETWEEN SCRAP CHARGING RATE AND OPEN HEARTH PRODUCTION RATE IN TONS PER HOUR



then graphically corrected to a constant value of Percent Melt Carbon and Scrap Charging Rate by a series of approximations. The resulting Tons per Hour values were plotted against the corresponding Percent Hot Metal values to give the curve shown in Figure 1.

These curves show that the effect of hot metal on tons per hour reaches an optimum at about 66 percent hot metal, and that the melt carbon should be between 0.60 and 0.80 percent. The effect of scrap charging rate on tons per hour is erratic.

There are two methods of modifying the multiple correlation method to test for interaction effects. These methods are explained in Chapters 21 and 22 of reference (3).

In the first method, the data are sorted in order of increasing values of one variable and then divided into two or more groups. Each group is then analyzed using the regular graphical multiple correlation method. Curves are obtained for each variable from each group of data. If the curves for a single variable do not have the same shape in each of the groups, interaction is present.

The data were divided into a low hot metal group (44.9% to 62.1% hot metal) and a high hot metal group (62.1% to 71.4% hot metal). The results obtained are shown in Figures 2 and 3.

In the low hot metal group of heats the optimum melt carbon is in the range between 0.80 and 0.90 percent, and fast scrap charging increases the production rate.

In the high hot metal group of heats the optimum melt carbon appears to be below 0.70 percent, and the rate of scrap charging is unimportant.

The second method of showing interaction effects involves the preparation of three dimensional graphs in which the (X) and (Y) axes are the two independent variables being tested. Values of the dependent variable are shown as contours. The resulting graph is similar in appearance to a topographical map, and may show slopes, peaks, valleys or ridges. Interaction is present when the contour lines are not parallel.

The following figures were prepared to illustrate this method. The dependent variable is production rate measured as heat time. The two periods of the heat, time before and time after the melt carbon test, were studied separately.

Figure 4 is a plot of Start Scrap Charge to Melt Carbon Test Time against Percent Hot Metal and Percent Melt Carbon. It shows that the first period of the heat is the shortest at high percentages of hot metal and melt carbon. However, the elimination of some carbon during this period of the heat is desirable and appears to cause the least delay when the hot metal is at approximately 65 percent. While high hot metal heats have the most carbon to eliminate to reach a given percent melt carbon, they also have the least steel scrap to melt.

Figure 5 is a plot of Time from Melt Carbon Test to Tap against Time from Start Scrap Charge to Melt Carbon Test and Percent Melt Carbon. It shows that the last period of the heat is almost independent of the time of the first period of the heat. The last period depends mainly on percent melt carbon and percent preliminary carbon, and is the shortest at

FIG 4. THE EFFECT OF PERCENT HOT METAL AND PERCENT MELT CARBON ON THE TIME FROM START CHARGE TO MELT CARBON TEST

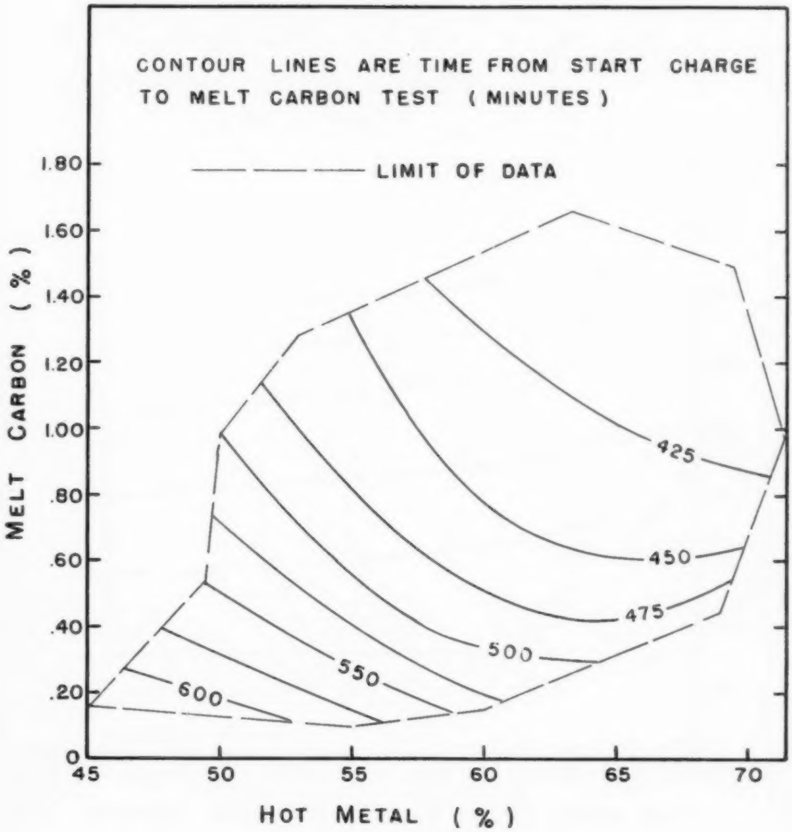


FIG 5. THE EFFECT OF PERCENT MELT CARBON AND TIME FROM START CHARGE TO MELT CARBON TEST ON TIME FROM MELT CARBON TO TAP

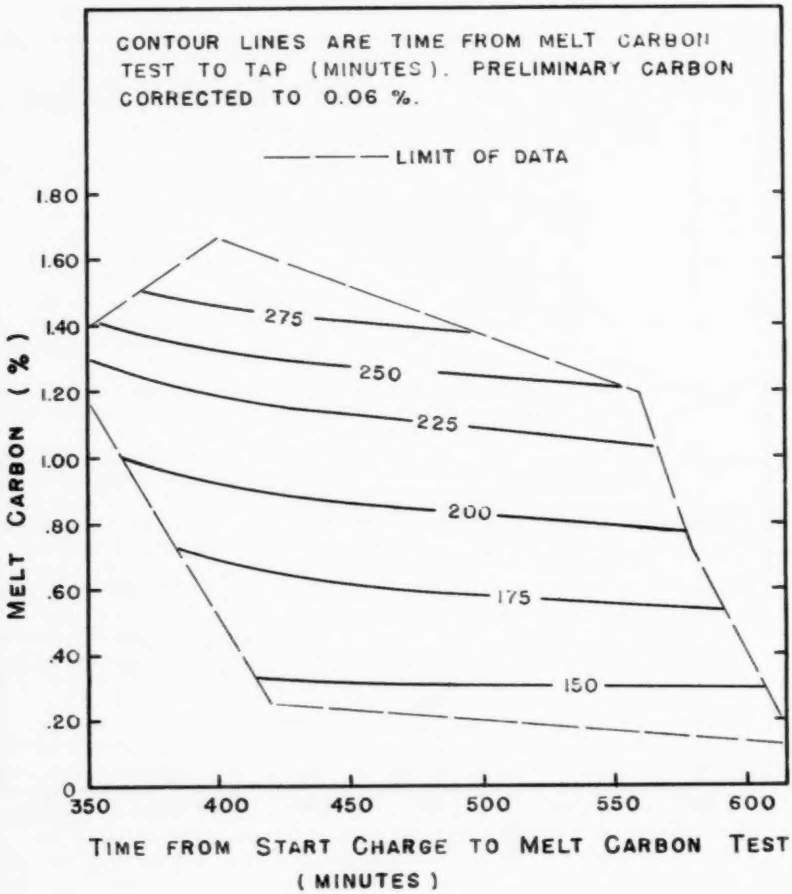
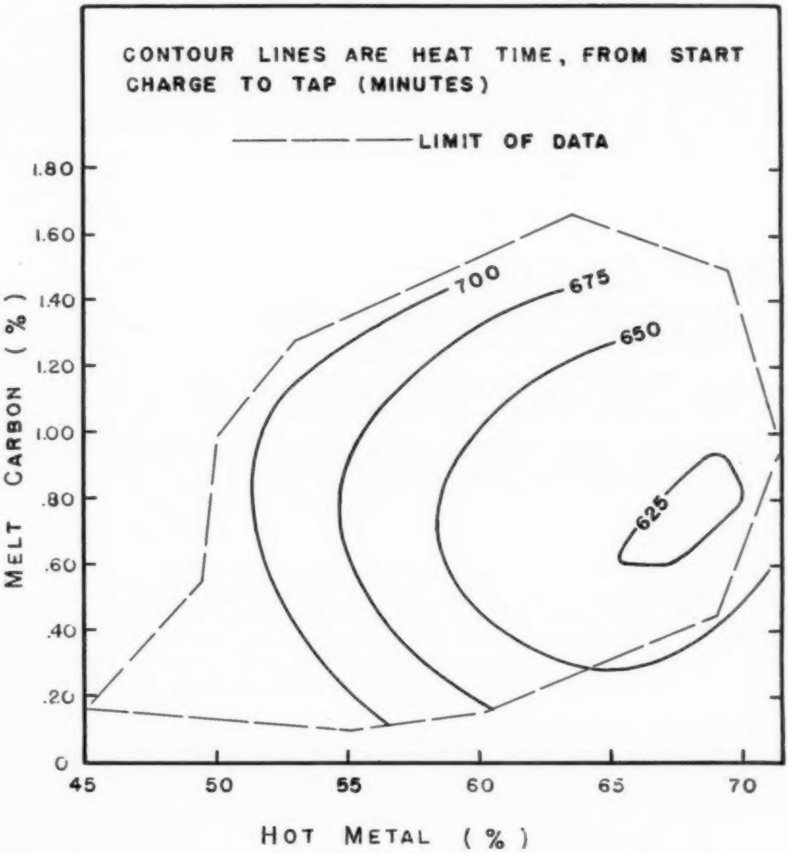


FIG 6. THE EFFECT OF PERCENT HOT METAL AND PERCENT MELT CARBON ON HEAT TIME



low percentages of melt carbon.

Figure 6 is a plot of Total Heat Time against Percent Hot Metal and Percent Melt Carbon. It was obtained by adding together the effects shown in Figure 4 and Figure 5. The conflicting effect of Percent Melt Carbon on delays in the two periods of the heat shows up in Figure 6 as a minimum point. For maximum production rate, the melt carbon should be in the range 0.60 to 0.90 percent, and the hot metal between 65 and 70 percent.

It is possible that interaction effects may be important in all types of industrial operations. They are known to be important in the making and processing of steel ingots.

The modified correlation methods allow the investigator to study these processes using either controlled or uncontrolled experiments, whichever may be most convenient. These investigations, when interpreted by experienced plant personnel, give management much of the information it needs to operate the plant most efficiently and expand its operations most effectively.

REFERENCES

1. G. W. Snedecor. "Statistical Methods", 1946.
   Collegiate Press, Inc.; Ames, Iowa

2. K. A. Brownlee. "Industrial Experimentation", 1947.
   Chemical Publishing Co., Inc.; Brooklyn, N. Y.

3. M. Ezekiel. "Methods of Correlation Analysis", Second edition, 1945.
   John Wiley and Sons, Inc.; New York, N. Y.

QUALITY CONTROL OF TUBULAR STEEL PRODUCTS

W. T. Rogers

While this paper is primarily concerned with quality control and tubular products made from steel, it is pointed out that the subject matter is applicable to quality control in basic industry generally and the material presented pertains equally as well to copper, brass, lead, glass, or an end product for customer consumption of any of the large industries which procure their own raw materials and perform all intermediate manufacturing steps. The title "quality control" may in itself appear restrictive in that it tends to limit the concept of the material herein presented to the accepted association of quality control with the use of control charts in conjunction with some form of sample inspection. These are useful tools and in all probability the main tools in any quality control organization, but in a basic industry such as a steel plant they must be supplemented by at least three other branches of statistical procedure, namely correlation analysis, experimental evaluation, and testing for significance.

It is not the purpose of this paper to criticize or to minimize the relative value of any statistical quality control procedure, but it is of importance to note that the concept of acceptance sampling of finished product for dimensional and visual defects has not been accepted generally by steel manufacturers or consumers. This does not mean that the use of sampling inspection is not appreciated because there are many instances of quality evaluation with reference to chemical, physical, and metallurgical characteristics which require some form of sample inspection. In addition, there are many points throughout the manufacturing process where quality is evaluated by means of periodic sampling procedures. These applications are continually being studied and applied by the majority of progressive steel plant operators.

It may be stated then, that there are two generally accepted concepts of statistical quality control, one of which is concerned only with acceptance-rejection sampling, and the other with control of the entire process and takes in almost the entire field of statistical methods. This second concept is the one that must be embraced in the quality control program of any large industry.

In order to project a clearer picture of the scope of activity associated with such a program it is necessary to sketch briefly the size and activity of the organization within which it is required to function. Lorain Works, of National Tube Division, United States Steel Corporation, employs approximately 12,000 people and produces in the neighborhood of 100,000 tons of buttweld and seamless pipe per month. The plant consists of a coke plant, docks and raw material storage facilities, 5 blast furnaces, 12 open-hearth furnaces, 3 bessemer converters, rolling mill and semi-finished product

---

W. T. Rogers is Product Engineer, National Tube Division, U. S. Steel Corporation, Lorain Works, Lorain, Ohio

conditioning facilities together with buttweld and seamless pipe producing units.

The statistical quality control activity may be classified in three sub-divisions all of which are integrated into the one common purpose of maintaining and improving the quality of finished products. These are as follows:

Process Control
Process Investigation
Process and Product Improvement

Process control is a routine procedure which consists of following those factors in each phase of production which have been found from past experience to be associated with the occurrence of unsuitable material. This part of the quality control program is supported by a system of statistical control charts pertaining to each department in the production system. This application makes it possible to keep each department head advised with respect to deviations from normal practice which are likely to result in higher than normal rejects at some subsequent processing point. Process control with respect to dimensional characteristics is a function of inspection and is likewise set up on a basis which lends itself to control chart analysis.

The second sub-division, process investigation, is a necessary adjunct of the quality control organization. This phase of the work consists of analyzing processing data for causes of abnormal departures from the limits established for any of the quality factors under observation. In statistical terminology this can be defined as the determination of "assignable causes". In most instances this is not a difficult task. However, there are times when extensive analytical work is required which the statistical quality control division is implemented to perform.

The third function of the statistical quality control group is to aid in the evaluation of changes in processing procedure and the development of new or improved products and processes. One of the accepted functions of management is the responsibility for improving practices in order to improve quality, increase production, and lower costs. In the highly complex business of steel production and fabrication each process has a multiplicity of factors related to these three objectives. It is therefore necessary to have some means of evaluation which will permit comparisons to be made on an equitable basis. These means are available in the various statistical techniques of correlation analysis, significance tests and analysis of variance and for this reason are a part of the statistical division's responsibility.

As in most steel plants, the statistical quality control group operates as a division of the Metallurgical Department and is directly responsible to the Chief Metallurgist at Lorain. It is staffed with 4 people; a statistical engineer with the status of supervisor, an assistant statistican, a junior statistican, and a statistical clerk.

The statistical engineer has a thorough knowledge of statistical procedures together with a broad background of mill experience. The assistant statistican has a working knowledge of statistical methods and extensive experience in the preparation of various statistical summaries and reports. The junior statistican is the coordinator between the Accounting Department, Tabulating Bureau and the statistical division and has a working knowledge of tabulating equipment and the preparation of tabulating reports and summaries. The statistical clerk receives and prepares incoming data for purposes of routine daily reports used in presenting daily quality control information to management.

The statistical division of the Metallurgical Department serves as a clearing center for quality control and product yield information among the operating, inspection, accounting, and metallurgical departments with the Chief Metallurgist making simultaneous reports with respect to abnormal conditions and recommended corrective action to top management and the operating departments.

The quality control procedures consist of the following operations:

1. Collection of data
2. Processing of data
3. Reporting quality control information to management
4. Maintaining quality control charts
5. Determination of assignable causes for abnormal conditions
6. Study of methods and practices for quality improvement

The collection of data is accomplished by the use of existing routine reports made out by metallurgical observers, chemical laboratory personnel, mill accounting clerks and the inspection department. Information is compiled on forms suitable to the purposes of all interested parties so that duplication of effort is avoided. Procedures are set up whereby the required reports are channeled through the statistical division. Practically all quality related data are punched on tabulating cards and wherever possible are collected in such a manner that they need only a minimum of checking in order to prepare them for the tabulating bureau. Eleven cards are punched for each heat of steel produced so that the entire history of the heat with respect to quality related factors is available on these cards.

The spark plug of the program is a daily quality control letter. This letter, which is forwarded to top management each day, contains a summary of inspection rejects for the preceding 24 hour period together with an indication of the state of control of each type of reject by class of material and producing unit. In addition, this letter also presents a condensed summary of quality related factors in each processing department with particular attention being called to those variables which show abnormal departures from the desired level of control. In addition to the daily quality letter, a summary of quality variables is prepared each week showing the comparative status of these variables for the current week, the preceding week,

and the most recent 6 months average. This is followed by a monthly summary which is prepared for discussion at a regular monthly steel quality committee meeting. This committee is composed of the various production department heads of the entire plant, the Chief Metallurgist, and Superintendent of Inspection.

The statistical division also maintains a set of some 200 control charts of which 20 are concerned with following the level of inspection rejects at various inspection locations throughout the plant. The remaining charts have the purpose of maintaining a check on the state of control of numerous quality related variables throughout the entire processing procedure. Figure 1 is an example of one such chart for following of the control of rejects and Figure 2 is a chart for a quality related variable.

The determination of assignable causes for departures from normal is also a function of the statistical group. Each heat of open-hearth and bessemer steel is identified from the start of its manufacture until the final inspection is completed and inspection results are summarized daily. When the daily average percent rejects exceeds the established upper control limit, the inspection results are examined in detail to determine the causes. When definitely assignable causes are found they are called to the attention of the processing department in which they occurred so that the department head may be apprised of the necessity for constant surveillance of factors within the processing area which contribute to the occurrence of inspection rejects.

In addition to its quality related functions, the statistical division also makes many special studies which could properly be classified as statistical research and which are concerned with the use of new practices and materials as they are related to quality and production rates. This function comprises a major proportion of the work load and is looked upon as equal in importance with the other statistical activities.

The basic approach to quality control of steel products is no different in the large steel plant than in the small metal fabricating shop. The same three steps of specification, production, and inspection apply here as in any other manufacturing activity. These three phases of quality control, while of similar importance in any industry, require a different approach with respect to the application of statistical methods. In small parts manufacture where defects occur at random and it is possible to secure random samples without incurring abnormal handling costs, statistical quality control is mainly concerned with the application of inspection sampling plans. In a large steel plant, however, which produces in the neighborhood of 100,000 tons of steel tubular products per month and where it has been established that abnormalities do not necessarily occur at random and where it is not economically practical to secure random samples of finished product, the statistical quality control approach must be different. The problem in this type of production is to control the various processing practices so that a minimum proportion of rejectable material will be delivered to the inspection areas.

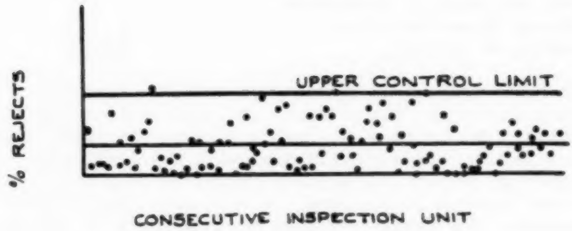


FIG. I CONTROL CHART FOR INSPECTION REJECTS

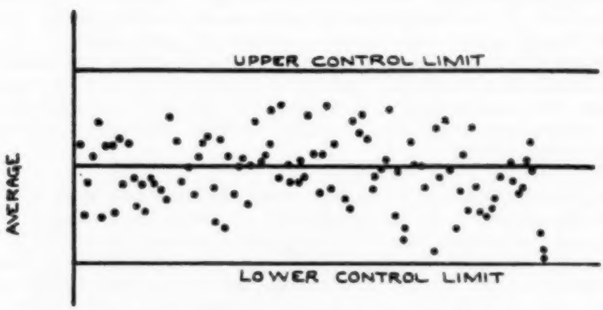


FIG. 2 CONTROL CHART FOR QUALITY RELATED VARIABLE

The problem of quality control in the manufacture of steel pipe and tubes can be classified under three headings as follows:

1. Chemical and physical properties
2. Dimensional characteristics
3. Elimination of defects

Some of the most rigid specifications in the line pipe and tubular products industry are those concerned with chemical and physical properties. The increased demand for higher strength material to support higher working pressures together with the chemical limitations imposed by present day welding practices require exceptionally close control of chemistry particularly with respect to carbon and manganese. The emphasis placed on this phase of quality control is not generally realized. It is a problem, however, in which statistical relationships and probability analysis have played an important part. For example let it be assumed that 0.33% maximum carbon and 1.30% maximum manganese on finished product have been specified together with a physical property which requires working to the top limit of both these elements. In steel making as in any other repetitive manufacturing process, no two heats are exactly identical in their chemical composition and so when working to a maximum specification limit it is necessary to aim for an average which will take the normal variability into consideration. The first samples for the determination of chemistry are taken from the molten metal as it is being poured from a large ladle into a series of ingot molds. At this point, the steel is in its most homogeneous state. Once the metal starts to solidify the physical laws of solidification and segregation tend to reduce the uniformity of distribution of chemical constituents so that any subsequent analysis is likely to vary somewhat from that of the original ladle tests.

It is apparent from the foregoing brief explanation that a knowledge of the relation of the chemistry of the molten metal, or what is known as ladle analysis in the steel industry, to that of finished product, or check analysis, plus the variability from heat to heat of steel, is essential in order to set up working limits which will meet the limitations of the customer's specification. This relation for carbon and manganese is presented in Figures 3 and 4, respectively. In Figure 3 it will be noted that with a range of .24% to .30% ladle carbon, the check carbon varied from .23% to .33% and not until ladle carbons of .30% are reached is there a possibility of exceeding .33% carbon on check analysis. It is therefore evident that a working practice which is unlikely to produce ladle carbon in excess of .30% will satisfy the customer's requirements. The range of .24% to .30% on ladle carbon indicates a standard deviation of .01% carbon which, with a normal distribution, makes it logical to assume that the aimed for carbon should be .27% or .30% minus 3 σ in order that practically no heats will exceed the specification maximum.

Figure 4 presents similar information with respect to manganese showing that with a range of .88% to 1.25% ladle manganese, a spread of .89% to 1.30% is obtained on check analysis indicating that to produce nothing over 1.30% on check analysis it is necessary to aim for

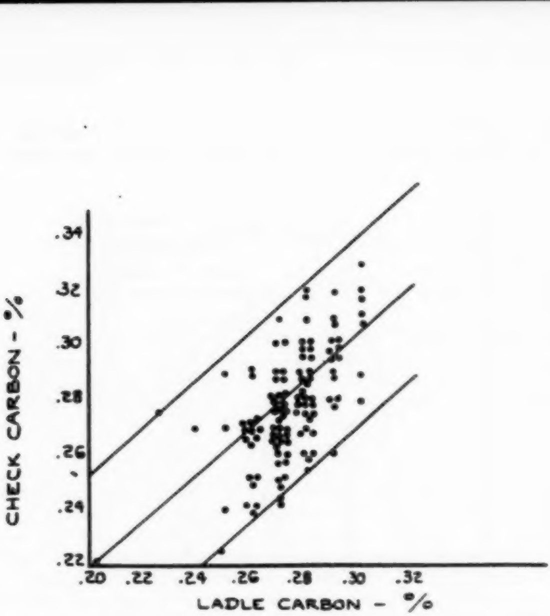


FIG. 3 RELATION BETWEEN LADLE CARBON AND CHECK CARBON

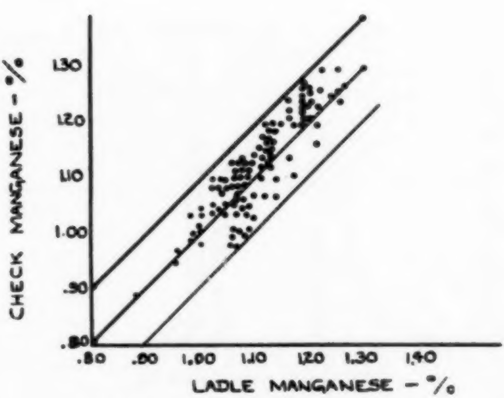


FIG. 4 RELATION BETWEEN LADLE MANGANESE AND CHECK MANGANESE

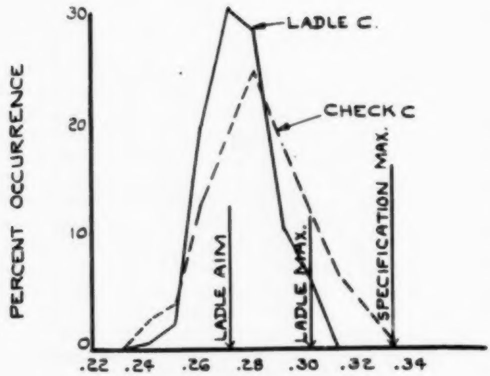


FIG. 5 FREQUENCY DISTRIBUTION OF CARBON

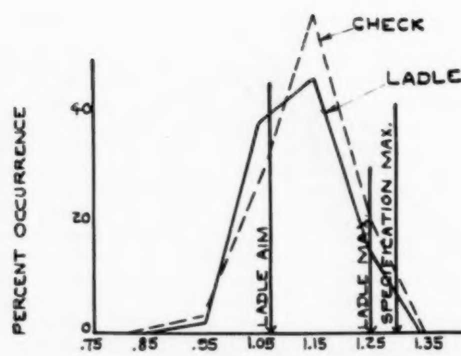


FIG. 6 FREQUENCY DISTRIBUTION OF MANGANESE

1.07% or 1.25% - 3 σ . On this basis the following comparison between customer specification and steel making aimed for analysis is obtained.

Customers Specification	Steel Making Aim Specification
.33 max carbon	.27 average carbon
1.30 max manganese	1.07 average manganese

The result of the application of this type of control is shown in Figures 5 and 6, where frequency distributions of check analysis are super-imposed on distributions of ladle analysis for both carbon and manganese indicating that the type of control expected was obtained.

Steel pipe and tubular products like other formed metal products are made to relatively close dimensional tolerances for such things as length, outside and inside diameter, weight per foot and wall thickness. These dimensional characteristics are subject to 100% inspection but in a great many instances a considerable lapse of time expires between the last forming operation and final inspection. It is therefore necessary to have some means of determining whether or not material being produced will meet the desired specification when it reaches the final inspection point. This is accomplished by a systematic sample inspection of the product immediately following the final production operation. By applying closer control than is actually required by the specification it is assured that finished product inspection will find practically all material well within the required specification tolerances.

Figure 7 is a control chart of groups of 4 measurements taken each hour on one of these dimensional characteristics. This chart indicates the statistical limits within which the material was produced together with the specification limits within which it could theoretically be produced and still have the total product within the acceptance limits. Figure 8 is a frequency distribution of this dimension on the entire lot represented by the Figure 7 control chart which shows that all material inspected conformed to the specification requirements.

The third phase of the statistical quality control program is that of investigating and studying various processing practices and procedures with respect to factors associated with the occurrence of defects in semi-finished and final product. In addition to the routine observance for the purpose of recognizing abnormal reject levels and the assignable causes for such conditions, effort is constantly being made to reduce the normal level of rejects. The statistical method used in this type of research is mainly multiple correlation analysis. It is realized that this method, like others, has its limitations but it has proved over a long period of time to be the most versatile, the easiest taught to non-technical personnel, the best adapted to punch card procedures, and the simplest to interpret to operating personnel of the various multiple factor analysis methods available. The use of this method as applied to a typical research problem is presented in

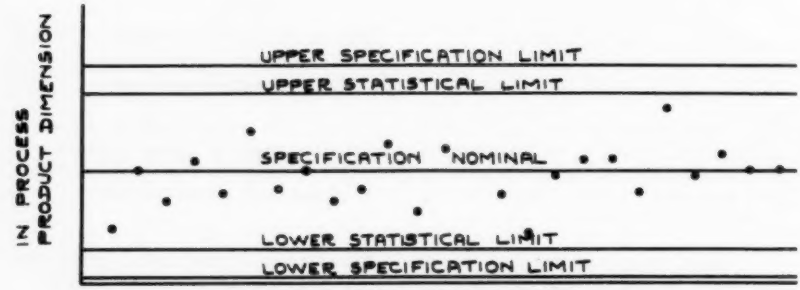


FIG. 7 - CONTROL CHART FOR DIMENSIONAL CHARACTERISTIC

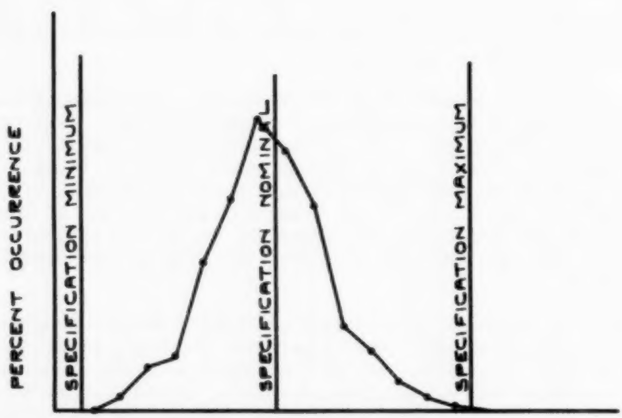


FIG. 8 - PRODUCT DIMENSION
FREQUENCY DISTRIBUTION OF FINISHED PRODUCT DIMENSION

the following example.

In the production of large seamless steel pipe, the first inspection opportunity is presented when the steel comes off the rolling mill in the form of a round billet. In general, the round billets are delivered to a peeling machine which takes a predetermined cut from the entire surface. This operation removes all minor surface defects and produces a surface on which more serious defects are readily discernible. The inspection personnel then decide which billets should be rejected as not suitable for further processing and which billets should be subjected to further conditioning. It is therefore evident that close attention to quality should be centered on this operation and effort is constantly being applied to reduce the percentage of material rejected at this point. This type of quality control therefore requires an extensive knowledge of the relation of processing variables to the occurrence of rejects.

One of the most important steps in this type of problem is the original approach. It is not a good practice for the variables included to be selected indiscriminately. The first step, therefore, is a detailed discussion of the problem with all interested parties. In this case, the approach to the problem was discussed with operating, inspection, and metallurgical personnel. As a result of this preliminary consultation it was decided that 31 factors should be examined with respect to their association with the occurrence of rejects. A period was then selected from an examination of the routine control charts during which abnormally high and abnormally low occurrences of rejects were observed. During the period chosen approximately 2300 open-hearth heats of steel of 160 ingot tons each were processed through this particular inspection area. Data collection was not a problem as the routine tabulating procedure employed kept all data on the various punched cards which were prepared daily. Card preparation consisted of transferring the numerical data for each heat to a working card and then, using this card, the necessary sums of squares and product sums for a multiple regression analysis were obtained directly from tabulating equipment.

Independent linear regression coefficients were then developed for each variable with the "index of rejects". The linear relations were then checked for curvilinearity with those having a significant relation being presented as shown in Figures 9 and 18. While these graphs are captioned in steel plant terminology and it is not the purpose of this paper to explain these terms, it is evident that some very definite conclusions are possible with respect to the relation of various factors to the occurrence of rejects in semi-finished steel. From this type of evaluation it is possible to select control levels for a great many quality related factors and attention can then be centered on the control of these variables knowing that such control will help in keeping rejects down to a minimum.

In concluding, it is pointed out that there are many more aspects of the complete quality control procedure than have been covered in this brief report. While some of these are not set up on a statistical basis, in most cases they lend themselves to statistical analysis when it is necessary to determine the causes for significant departures from normal processing practice. In addition, the

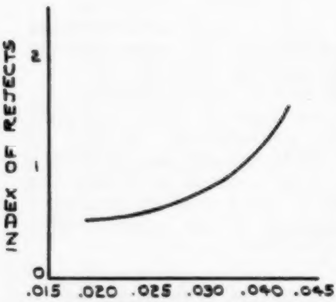


FIG. 9 - PERCENT SULPHUR

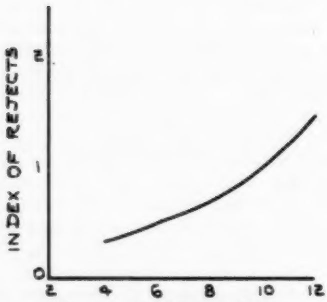


FIG. 10 - PRODUCT SIZE

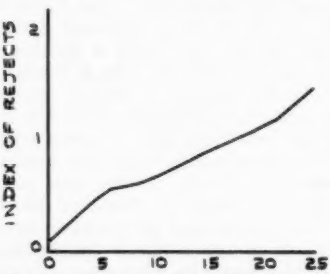


FIG. 11 - SKULL WEIGHT/1000 POUNDS

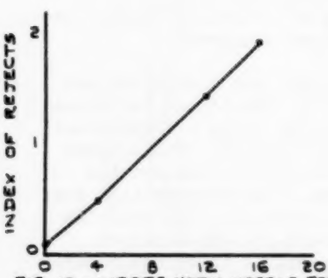


FIG. 12 - INGOTS WITH NOZZLE TROUBLE

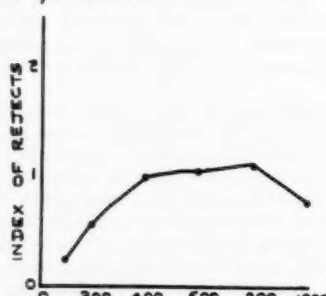


FIG. 13 - INGOT DELIVERY TIME IN MINUTES

RELATION OF STEEL MAKING FACTORS TO INSPECTION REJECTS ON SEMI-FINISHED PRODUCT

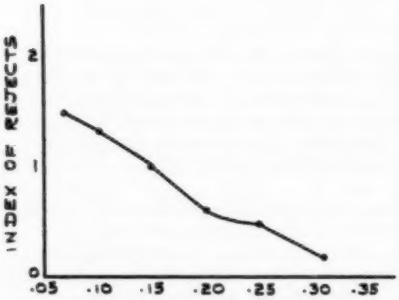


FIG. 14 - PERCENT TAP CARBON

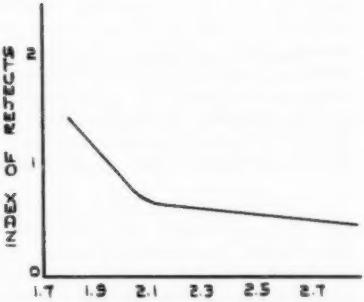


FIG. 15 - LIME/SILICA RATIO

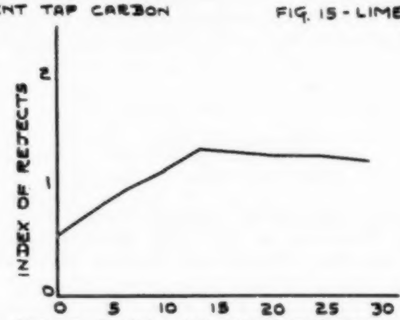


FIG. 16 - ALUMINUM AS MOLD ADDITION
OUNCES PER INGOT

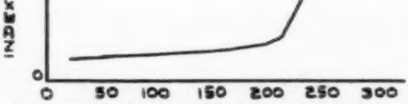


FIG. 17 - REFINING TIME IN MINUTES

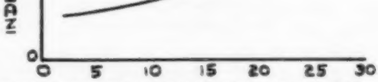


FIG. 18 - TAPPING TIME IN MINUTES

RELATION OF STEEL MAKING FACTORS TO INSPECTION
REJECTS ON SEMI-FINISHED PRODUCT

functions of the statistical quality control group include the application of various other techniques in determining the significance of changes in practices which develop from the introduction of new equipment, different raw materials and improved methods. Some of these other methods can be classified as probability analysis, moving averages, trend analysis and significance tests. Lorain Works of National Tube Division of United States Steel Corporation has been a pioneer in the introduction and application of the science of statistical analysis to the many problems which are associated with quality control in the steel industry as indicated by the appended list of references of published material.

It may be concluded from present usage and development that the use of statistical methods in connection with quality control procedures is generally accepted by basic industry. The employment of these methods in this type of industry have a much wider range of application in the fields of process control and investigation however, than in the otherwise accepted areas of sample inspection which have proved to be suited to the lighter fabricating industries.

References

1. "Application of Statistical Methods to the Solution of Metallurgical Problems in the Steel Plant", W. C. Chancelor. Proceedings American Society for Testing Materials, Vol. 34 Part II 1934.

2. "The Utility of Statistical Methods in Steel Plants", H. J. Hand, Transactions American Institute of Mining and Metallurgical Engineers, Vol. 131 1938.

3. "The Practical Application of Statistical Methods in a Quality Control Program", W. T. Rogers, Transactions American Society for Metals, Vol. 25, 1945.

4. "Multiple Correlation Applied to Steel Plant Problems", W. T. Rogers, Transactions American Society for Metals, Vol. 37 1947.

5. "Operation of Statistical Quality Control in a Steel Mill", W. T. Rogers, Steel, March 14, 1949.

6. "Statistical Analysis of the Effect of Alloying Elements on Mechanical Properties of Seamless Steel Tubes", W. T. Rogers, Transactions American Society for Metals, Vol. 40, 1950.

7. "Use of Oxygen in the Bessemer Converter", W. T. Rogers and L. T. Sanchez, Journal of Metals, September 1952.

CONTROL OF QUALITY OF BRASS STRIP

L. W. Thelin - Quality Supervisor
Chase Brass & Copper Co.

Some two or three years ago, our good friend, and a past President of the American Society for Quality Control, Mr. Samuel Collier, visited our home plant in Waterbury, Conn. The purpose of his visit was to look over the program we had under way to maintain and improve the quality of our products. I remember many comments he made during that day with us, but particularly I recall his saying words to this effect, "Use a plan which best fits your type of operation." In order to maintain a competitive position today, our mills must look to the quality of our products. I am sure that in the past "quality" was a responsibility of many of our supervisory and production people. These included Engineering, Metallurgical, and Inspection personnel, as well as the production people themselves. A number of years ago it was decided totake a more definite action, developing and putting into operation a quality group operating within the Metallurgical Department. (SLIDE) Our group is relatively small, and this slide shows our line of authority extending back up to the Works Manager. We originally concentrated on our Sheet Mill operations. You see three Strip and Sheet Inspectors who were brought into this group. Each of our individual mills, Tube, Rod, Wire, and Sheet, has a Quality Inspector assigned to them, and recently we added a second and third shift Quality Inspector to broaden our coverage around the clock. Our experience has shown that if our standard operations are maintained properly the quality of our product is satisfactory.

Bear in mind, we are essentially a reasonably large jobbing shop making large and small runs of copper and brass alloys in a number of finished and semi-finished commodities. Our finished mill products include brass strip and sheet in widths up to 24", the majority being shipped in coils of a considerable range of weights per inch of width. Our standard bar is approximately 30 pounds to the inch of width, and the weight can be increased to 120-150 pounds per inch of width by welding coils together. Welding is accomplished with our common alloys by using the heliarc method.

Copper and brass tubes in straight lengths and coils are made in sizes from approximately 1/4" to 14", and, of course, in a wide range of wall thickness.

We also fabricate brass rod and wire, supplying the entire range of sizes from pin wire 0.010" diameter to 5" round rod. Our Rod Mill production also includes a very large volume of extruded architectural and commercial shapes, as well as a sizable quantity of welding rod alloys.

The standard commodities referred to above are made in some 25 common or standard brass alloys, and possibly 15 or 20 not so common. This makes a total of 40 to 45 different alloys that are cast and run through our mills, in addition to two classes or types of copper in sheet or strip form, and three or four classifications of copper furnished in tubular items.

In the case of our Strip Mill, all alloys are cast in Ajax-Wyatt type low frequency induction furnaces of approximately 3400-pounds capacity. Using this type furnace for melting, and with a proper charcoal cover on the melt, oxidation is held to a minimum and thorough mixing of the molten metal takes place due to the pinching action in the furnace. Flat bars are usually cast 25-1/2" wide a 2-1/4" thick, 60" long after shearing. These bars have a cast weight of 1000 to 1100 pounds. Standard size extrusion billets are cast 8" in diameter and 100" long. These castings (billets) weigh 1500 pounds each, and are the starting form of all our rod and some brass tube products.

Early in our program we encountered a siege of surface defect troubles which caused relatively large rejections in the shop, and of course our Metallurgical staff was on the spot. Extensive controls were established on all three shifts for a considerable period. Mold dressings and the mode of applying dressing, as well as pouring techniques, were carefully gone over but no one cause could be determined which had produced the difficulty. A particular defect is well recognized in the brass industry, and is called a "spill". (SLIDE) Typical or characteristic spills are shown in this slide. It is generally conceded that slow pouring or cold pouring temperatures will produce defects in the casting which, when rolled out, become spills. However, even under the best controlled casting procedures a small percentage of defects is produced, and only constant care in the mill at points of process inspection can keep this trouble at a livable minimum.

Our experience to date has indicated that this particular defect most likely originates from unsoundness just below the surface of the bar, and from splashes or cold shots (called shotty bottoms) at the bottom of the castings.

In order to send to the mill the best possible castings, we decided to incorporate in our gate shearing operation a surface and soundness inspection. To do this, it was necessary to design a mechanism for turning over the cast bar weighing about 1100 pounds. The Mathews Conveyor Company of Elmwood City, Penna. was consulted, and here is the device they designed, and which we had installed in the line. (SLIDE). With the installation of this procedure our daily final inspection showed a marked drop in the so-called spilly metal rejects in our Finishing Department.

We can only assume this improvement was due to the fact that the under surface which, in the past, had gone down through the mill without any inspection was now being examined and defective units removed. There is now no opportunity to cover up poor castings by turning the poor side down in the stack.

For some years now we have maintained what we call 100% analytical control of all the castings we make. Improvements in chemical procedures, photometric analysis, and rapid spectroscopy have made it possible to analyze each and every round of metal cast. With allowances for experimental errors in testing and errors of sampling, we can be sure our material is within the chemical specifications before it is released to the mill. By so doing, we have eliminated one variable, and have assured ourselves that all our material going to our customers is of the proper mixture.

One of our most recent additions to our analytical control equipment is our Quantometer which enables us to analyze quantitatively up to eighteen elements simultaneously. This equipment (SLIDE) which is extremely rapid makes these determinations in less than two minutes. The advent of such a device has brought about a complete change in the analytical control procedures for our Casting Shop.

Earlier methods by electrolysis required milling or drilling of cast ladle samples, dissolution in acid, and plating on platinum electrodes (method for copper and lead only). This method is still used for miscellaneous testing of fabricated samples which are not usually in a proper form for direct-reading spectroscopic analysis.

Briefly, the operation on the Quantometer consists of sparking a freshly machined surface on the cast ladle sample. The luminous vapors formed by the spark (1000 volts uni-directional) are analyzed by a grating spectrometer, and the energies of the respective element lines are recorded and amplified by a system of exit slits, photomultiplyer tubes, and condensers. The final stored voltage is recorded on a Leeds & Northrup strip recorder which has been calibrated for the material we are working with.

This unit is manned by two people per shift, and has resulted in a considerable manpower saving in our control laboratories.

With such modern improvements we can be sure all of our material is within the proper chemical specification before it is released to our mills for fabrication. (SLIDE) - <u>Distribution</u> - <u>Copper Analyses</u>. Av. 61.32.

At the same time, we have established a closer control of inspection at our overhauler operation, so that bars containing surface defects which did not clean up, i.e., removed by overhauling, would either be put through the milling machine a second time, or would be scrapped if the surface was still questionable after the second overhauling. (SLIDE) - <u>Overhauler</u>.

After rolling on our 15" mill, which operation follows the overhauling, the bars are in coil form and it becomes increasingly difficult for a process inspection to do much about physical defects or blemishes on the coil surfaces.

The maintenance of uniform gauge through the intermediate rolling sizes is probably our Number One problem. This is evidenced by the fact that gauge variation has been a source of difficulty on numerous occasions. Any of you who may be connected with rolling mill activities know that the present day demands for close tolerance work necessitate closer controls and greater care in the actual rolling operations.

With our semi-continuous conveyorized handling of large coils (1000 pounds or more) hand gauging with micrometers is, of course, not feasible, and we have to use some automatic or "in the line" device to maintain gauge. For several years we have used the X-ray tube gauge. X-ray gauges are installed on all of our rolling mills, starting with our 15" mill through our tandem mills, and down to the finishing 8-1/2" and Bliss 2-high mills.

Our X-ray gauges are built by one of the industrial gauge manufacturers. The gauging head consists of the X-ray source enclosed in a ray-proof chamber with proper size diaphragms for specific purposes. An incident tube which converts the X-ray signals to voltage variations is on an arm of suitable length. The incident tube is spatially related to the X-ray source so that strip to be measured passes untouched through any portion of the gap between these units.

The signal from the gauge head is led into the electronic circuit where it is amplified, converted to a linearly varying voltage, and fed into an output circuit where it increases the zero setting voltage, and so operates the thickness deviation indicator.

The theory of the X-ray as a gauge, of course, is due to the characteristic adsorption of X-rays by the various metals (the adsorption being determined by the density of the metal). The X-ray gauge measures the difference in energy adsorbed due to lighter or heavier metal passing through the gauge.

The control panel is mounted where it is accessible to the operator and easily read by him. Zero settings are made by using carefully prepared standards of various alloys to be rolled, and of carefully selected thicknesses. Accurate gauge can be maintained at all reasonable speeds of rolling. The use of the X-ray gauge with automatic screwdown adjustment plus manual adjustment has allowed us to meet increasing demands for close tolerance in strip.

Our production of copper and brass strip is, in general, divided into two categories insofar as temper is concerned. One is cold rolled or roll tempered, and the other is annealed strip which is suitable for further cold working.

From a quality standpoint, we do not have much of a problem with rolled tempers - a wide range of hardnesses can be obtained, and the hardness can be held to quite close Rockwell hardness numbers. The widespread use of the Rockwell Hardness Tester by both mill and customers has eliminated guesswork and objectionable variation in this class of materials.

Annealed material, however, presents a somewhat different problem. Considerable improvement has been made in mill annealing equipment in the past ten years. With the exception of our first breakdown annealing, which is done in roller hearth oil-fired furnaces, most annealing units are gas-fired, (They have recently been changed over to natural gas), and are supplied with a protective atmosphere to minimize scaling or oxidation. A recent improvement in copper strip annealing has been the use of the bell type annealers which furnish bright annealed material not requiring pickling or cleaning of any kind. Copper and high copper content brasses are handled in these units. The load is started on a circular carrier which is lifted onto a circular hearth, then is covered and sealed by a water and inner oil seal. Air is purged from the bell by passing dry inert gas into it for one to one and one-half hours. When oxygen is removed, the bell furnace is lowered over the bell and the temperature brought up to the required point. When the proper temperature is reached, the charge is allowed to soak for the correct time to complete the anneal. The load is allowed to cool to a low temperature before opening the bell so as to prevent staining or

oxidation. Cooling is facilitated by spraying the outside of the bell with water.

Bright annealing furnishes a cleaner, brighter surface finish than available heretofore, and as such constitutes a real improvement in the quality of commercially obtainable copper and copper alloy strip and sheet.

In addition to many improvements in annealing furnace construction and operation, such as controlled atmosphere, continuous operation, better uniformity of heating, a recent step forward in the quality of annealing has been made by the development of continuous strand or single strip annealing using electrical heating.

Single strand annealers are available which take the full 24" width and pass it through a heating zone at speeds producing 6000 to 8000 pounds an hour. Under proper operation, these furnaces produce in tonnage quantities of the best uniformity of anneal of any production units. Continuous operation is maintained by stitching coils together (mechanical joints). After the anneal the strip is passed through cleaning acid, followed by rinsing and drying. The finished annealed strip is of uniform color, and free of so-called annealing scallops generally associated with earliest methods of annealing and cleaning of brass strip.

Several years ago we were considerably handicapped by inaccurate data from our inspection personnel. Consequently, it was decided to take the responsibility for inspection of strip and sheet materials out of the production group and place it where it more properly belongs, with the Quality Control group. Early in this period, we were surprised ourselves to find a rather large number of conditions which warranted rejection or reworking of strip material. A revised plan of inspection reporting was developed, and complete records kept from then on. Daily summaries were sent to the Mill Superintendent. This resulted in immediate action on any operating condition by the production group. More deep-seated troubles were assigned to the Quality Control group, as well as to the Metallurgical Engineer, for complete investigation. Some problems were corrected very promptly, whereas others may require much longer range effort to obtain the solution. Still others may require the co-operation of the Research and Development people to effect a proper and permanent solution.

Our inspection procedure is set up so that the line inspectors, who are part of the finishing crews, report to the Shift Supervisor in charge of inspection, and he, in turn, reports to the Quality Engineer.

The primary function of the Inspector at this period is to check the material for gauge, surface, and all other characteristics which may be critical in accordance with the customer's orders or specifications, or both. With respect to gauge, all orders may be classified into two groups:

A) Commercial gauge, on which standard specifications or Copper & Brass Research Association (CABRA) tolerances apply.

B) Closer than commercial tolerances.

In the first case, the coils or bars are gauged according to Product time standards. Generally speaking, this means gauging in three places on each coil. The starting end is gauged and cut back until the bar is on gauge, and the slitter stopped once at the center of the bar for gauge checking. The end of the bar is again cut back if found off-gauge.

Since we have no difficulty meeting the requirements in the first case, most of our effort is put on the second, or closer than commercial tolerance, classification. For orders of this description, we clearly mark "SPECIAL TOLERANCE" on the face of our mill ticket, and no material can be passed which falls outside the specified dimensions for gauge. We felt that this category would lend itself to checking on a statistical basis, and we have worked out a sampling plan which is now in effect. Sample sizes of two to fifteen bars are checked, depending on the total number of bars in the lot. The sample bars are gauged manually in four places through the bar. At each gauge stop a length of 8 ft. is gauged at 1 ft. intervals. If the sample bars are found to be satisfactory, i.e., on gauge, the remainder of the lot is run on current standard. If the sample number is found defective, i.e., off-gauge, all bars in the lot are put through this tightened inspection procedure. I might say that many of our Inspectors have worked so long by earlier methods that it may be too soon to evaluate clearly this newer approach to a mill quality problem of long standing.

Even a short talk would not be complete without some reference to surface quality. We found just as soon as we got into this problem of inspection of strip and sheet that much or our trouble stemmed from loosely defined standards, and multiple interpretations of some of the standards. We recognized the fact that an increasing percentage of our orders was coming from our customers marked by such terms as "Polishing Quality", "For Chrome Finishing", or "To Be Used for Jewelry", and so forth. It was necessary then to make an acceptable and real definition of surface finishes, and have all our Inspectors abide by these differentiations. A careful survey of our orders showed that they could be grouped into four definite classes as regards surface finishes. These we have designated - Etching Quality, Jewelry Quality, Polishing Quality, and Commercial Quality. Our definitions of these are as follows:

<u>Etching Quality</u> - this applies to the highest level of quality in strip used for etched metal dials, name plates, and similar articles. No visible defects of any kind are permissible in this type material.

<u>Jewelry Quality</u> - this applies to strip used mainly by the jewelry trade, but includes engravers brass and reflectors. This classification permits slight shape defects, and slight discolorations or stains which can be readily removed by acid dipping or bright color buffing only; but does not permit any casting, mechanical, or surface textural defects whose removal would require substantial buffing, or other, more drastic, surface treatment.

<u>Polishing Quality</u> - this applies to strip for hardware, plumbing lines, mirrors, and similar operations, and permits slight defects of all types which are removable by ordinary cut and color buffing.

Commercial Quality - this applies to all strip and sheet in the absence of the above quality designations. This standard permits slight or moderate defects of all types indicated above, but which are removable by grinding or emerying, followed by cut and color buffing.

Aside from gauge and surface considerations, our Strip Mill Inspection Manual lists some ten characteristics applying to dimension and shape; six or seven conditions applying to annealing and cleaning; three or four defects stemming from the Casting Shop; as well as fourteen or fifteen recognizable defects due to handling in the mill.

With this considerable number of potential trouble spots which may mar a previous good quality record, it is not difficult to see why people responsible for final outgoing quality may some time or other feel that everyone in the plant is against them, and that they see more questionable metal quality than good.

A serious effort has been made in working out the training of personnel to be Quality conscious. Our customers are becoming more and more quality minded. Refinements are being added to many items which in the past we took for granted, and consequently today we must supply larger, flatter, cleaner, and more accurately rolled strip than ever before. This has necessitated a thorough understanding of many of our customer requirements.

Manuals for the aid of Inspectors and Supervisors have been prepared and, more important, are being used so as to minimize guesswork and to aid in making decisions.

In addition to a continual poster program largely quality motivated, we have made a careful survey of our Inspectors' eyes since we rely heavily on visual inspection. Eye ratings have been made on about 75 people using an Ortho-Rater which machine determines various eye characteristics such as color and depth perception and muscular balance, all of which are important to anyone charged with the responsibility of visually examining material of any kind.

While our actual staff of people formally recognized as the Quality Control Group is not large, we have extended our coverage to all shifts and we have attempted to tie down the critical points, qualitywise, in our operation. In this manner, we are producing the best quality strip we know how consistent with our economical overall operation.

L. W. THELIN
QUALITY SUPERVISOR
CHASE BRASS & COPPER CO.

Kennecott-Chase movie, "Science of Making Brass", is to follow this talk.

DEPARTMENT STORES USES OF STATISTICAL QUALITY CONTROL*

Claude S. Brinegar
The Emporium, San Francisco

The modern, large department store is deceptive. In appearance, its main activities seem to involve only the selling of merchandise to customers. In fact, however, for every employee assigned to the selling floor there are probably two employees engaged in nonselling activities. And for every piece of paper needed for a customer transaction, there may be five pieces needed to get the merchandise from the vendor to the selling floor. Such diverse activities mean, consequently, that the majority of Statistical Quality Control opportunities are found not on the selling floor but behind the scenes.

These behind-the-scene applications are found in the accounting departments, in the operational division, and in different aspects of overall management. I will discuss some interesting examples of such applications, drawing upon experiments carried out at The Emporium, San Francisco's largest department store.

A good illustration of an accounting application is found in Accounts Payable, the department that pays the bills owed <u>by</u> the store. At The Emporium we suspected that completely double-checking the invoice extensions made by three experienced comptometer operators was unprofitable, although some control was desired. Analysis of 15,000 invoices demonstrated (1) that costly errors were concentrated in the invoice representing very large purchases and (2) only very few invoices actually contained errors. This situation clearly suggested the possibility of a sampling plan of the acceptance/rectification type. We inaugurated such a plan, treating each operator's output of about 1,000 invoices under $500 in value as an inspection lot (the larger-valued invoices continued to be double checked). The inspection lots, which constituted about 80% of the total, were sampled when accumulated, usually once a day, according to a previously designated Dodge-Romig double sampling AOQL plan. If the sample was acceptable, the entire inspection lot was accepted as containing correctly extended invoices. If the sample was rejected, the operator was notified and the entire inspection lot was re-extended. Since the operators were capable of producing work well within the AOQL used, and usually did, the principal causes of errors were carelessness, fatigue, or illness. Significantly, Mondays and Fridays were the days most likely to see rejected lots.

This plan enabled us to eliminate nearly eight hours of checking-time daily, without materially increasing the number of undetected errors. Subsequent analysis by auditing teams indicated that less than 1/10 of 1% of the accepted invoices contained errors. In addition, since the operators soon materially improved their accuracy, we were able to lighten the

---

*The author wishes to acknowledge his gratitude to the executives of The Emporium for their cooperation, advice, and assistance. Messrs. Howard Carver and Walter Kaplan deserve particular mention. Miss Betty Legarra of the Research Department assisted materially in carrying out the actual experiments. The author is currently employed by the Union Oil Company of California.

inspection standards in a short time, producing additional savings in time and expense. New or inexperienced operators were sampled with the stricter plan until their abilities were established.

A similar application was found in Accounts Receivable, the department handling bills owed to the store. The charge account slips are filed in large, alphabetically-arranged drawers, with each drawer containing the file cards of about one thousand accounts. The drawers are assigned to individual clerks, who are then each responsible for correctly filing about 20,000 of the charge slips monthly. Since misfiled slips usually result in billing one customer for items not purchased, and submitting a corrected bill to the other customer, such errors produce adverse customer reaction, and, of course, should be minimized.

The established procedure for detecting the misfiled slips required each of the twenty clerks to devote about one hour daily to checking the files of one of the other clerks. The checking was such a repititious procedure, however, that even after 100% verification, we still received about 50 customer complaints monthly that were directly traceable to misfiled slips.

We started keeping detailed records of all misfiled slips located by the checking, and within two months found that three of the twenty clerks were responsible for nearly 90% of all errors. Under close supervision these three clerks soon became as efficient as the others, for they had apparently simply become careless. We instituted a sampling plan similar to the acceptance/rectification plan described previously--using one drawer as an inspection lot--and within a few weeks practically no errors were being detected. At the present time only the filing of new clerks is double checked, with the work of the regular clerks sampled at infrequent intervals. This application enabled us to nearly eliminate the checking, while, at the same time, the number of customer complaints traceable to misfiled tags remained essentially unchanged.

Similar illustrations exist in other accounting departments, such as Sales Audit, Internal Audit and Payroll. However, since these examples are analogous to the two just described, I will not go into details. Instead, I would like to turn to some interesting applications found on the operating side of the department store, describing experiments both at The Emporium and elsewhere.

An interesting application was reported recently by a large Eastern store. Their wrapping and packing department was experiencing an unusual number of complaints due to damaged goods or improperly filled orders. Since the orders could not be checked completely without undue expense, the store decided to install a sequential sampling scheme to control the quality of the wrapping and packing work. Samples were drawn at random time intervals, and when low quality work was detected, subsequent packages were completely checked until a specified number were found to be correctly processed. The plan proved to be quite successful, with the employees becoming quality conscious and customer complaints quickly dropping to reasonable levels.

The J. L. Hudson Company, Detroit's largest and the nation's second largest department store, has applied sampling techniques to study the activities of their supervisory personnel. By analyzing the actions of their 150 floor managers at random intervals during the day--they call if "work sampling"--the management was able to fit together an accurate

picture, at a small cost, of just how the floor managers spent their working hours. An interesting finding--and one that all customers must suspect occasionally of all stores--was that their managers had so much supervisory and detail work that they were forced to neglect their primary responsibility of attending to customer's needs. Consequently, in addition to substantially increasing their staff of floor managers, the store also divided the job into different responsibilities, thus providing much-improved service to their customers. The "work sampling" studies are being extended to appraise other important supervisory jobs.

We have used sampling for similar purposes at The Emporium. For example, by analyzing the sales pattern of randomly selected departments, we were able to determine, from a 10% sample, just how the sales were distributed throughout the day. This information enabled us to evaluate the expenses of night openings and other alternative operating schedules.

Statistical sampling has also been used to advantage in analyzing the competitive effects of branch store operations. By studying the shopping habits of a randomly selected sample of charge account customers, we were able to obtain an approximate measure of the extent to which The Emporium's new, nearby branch store was cutting into the sales of the main store.

The examples discussed thusfar were selected principally to illustrate the variety of department store operations that can be improved by the use of Quality Control techniques. The sampling plans and other details have been discussed only broadly, for I believe we are interested mainly in the kinds of opportunities that exist. There is, however, one application that I would like to consider in greater detail. This application, which involves the use of Quality Control in taking the merchandise inventory, is of general interest not only because the inventory is an essential but costly part of operating a department store, but also because the technique we have employed has potential uses in many other businesses where merchandise inventories must be physically verified from time to time.

The merchandise inventory is taken in two parts, the reserve (warehouse) stock and the forward (on-the-floor) stock. The count of the reserve stock presents no unusual problems. However, the count of the forward stock, which may include as many as two million separate items, is complicated by the requirement that no merchandise ready for sale should be tied up more than a few hours. The inventory is usually taken annually, as at The Emporum, and involves nearly every executive and regular employee in the store. It is, of course, necessary that the count be quite accurate, for the inventory figure is used for many accounting purposes, including the determination of total profits.

The forward inventory count is carried out by dividing each department of the store into sections and assigning a two-person team to each section. These teams may consist of regular store employees or temporary workers hired especially for the inventory. One member of the team calls information about the stock from the shelves, tables, bins, etc., and the other enters the items on the inventory sheet. The stock is listed by price, description, quantity, classification within the store, and season letter. This count is carried out under considerable time pressure. Prior to our investigation, a 100% check had always been made of all work performed in the forward inventory, as it was believed that such a check was necessary for reliable results. This 100% check nearly

doubled the cost of the inventory and also nearly doubled the time required.

We analyzed the existing procedure to determine the possibility of substituting an acceptance/rectification sampling scheme for the 100% double check of the forward inventory. Such a procedure seemed feasible, since the errors that were presumably corrected in the 100% verification were those within the control of the inventory teams. Fortunately, information about the types, frequency, and severity of errors that had been discovered by the check inventory teams in past years could be obtained by examining the old inventory sheets. No corrections to inventory sheets have ever been permitted by erasure; every correction requires calling the inventory supervisor, crossing out the improper entry, and making the corrected entry at the bottom of the sheet. By examining the handwriting of the corrected entries, we could tell whether the correction had been made by the original team or by the checking team.

Some 3,000 inventory sheets from the previous year were analyzed in this manner. The analysis indicated clearly that a few teams were usually responsible for most of the errors made in any department. A large majority of the teams produced work of quite respectable accuracy. Large errors in price or quantity were infrequent, the most common errors being over or under one unit in physical count, one dollar off in price, or listing the wrong season letter. Different types of goods seemed to differ greatly in their liability to inventory error.

As a result of this analysis, we inaugurated an inventory sampling plan for the January 1950 inventory. (The 100% double check was maintained in a very few spots, particularly on lines of merchandise where an error in count would prove costly). The inventory took place during two 4-hour periods from 6 to 10 p.m. following regular working days. The work of each inventory team was sampled by a "flying inspection squad" of experienced store employees. If the sample proved satisfactory, the entire night's output of the team was approved. If the sample contained too many errors, the team was immediately notified of the fact and its procedure was closely observed; all of its previous work was checked 100%; subsequent work was sampled again (on a stricter basis) with the possibility of a future 100% check.

For the sampling procedure, we selected three different Dodge-Romig double sampling AOQL plans. The plan to be applied to each department was selected on the basis of the analysis of the previous year's inventory sheets, giving weight to the type and value of merchandise and the probable number of inventory entries. In the use of the sampling tables, one line on the inventory sheet (typically representing 10 to 125 items) corresponded to a single manufactured article subject to inspection; any error on the line caused its classification as a defective. However, a line that was correct except for an error in season letter was counted as only one-half a defective; this error affected neither the count nor the value of the merchandise, but only its estimated age distribution.

The sampling plans were selected having OC curves indicating that not more than 20% of the total work would be rejected and reinspected. This decision was an essential part of the planning of the inventory as it determined the number of reinspection teams to be provided.

The result of this inspection scheme for three different years is summarized in the following tabulation, showing for each year the number

of teams inspected and the percent rejected:

Year	Number of Teams Inspected	Percent Rejected
1950	462	10.8
1951	463	12.5
1952	535	11.0
Total	1,460	11.6

For each of these three years the rejection rate varied within narrow limits, from a high of 12.5% to a low of 10.8%. The sampling procedure functioned smoothly, with no more than the anticipated number of difficulties. The substantial increase in teams inspected in 1952 was the result of adding departments that were previously checked 100%. As we became increasingly confident in the ability of the "flying squads" to single out the inefficient, error-prone teams, we 100% checked fewer and fewer of the departments containing expensive price lines.

Near the end of each inventory there were invariably a few idle teams. All inventory supervisors were instructed to use these idle teams in double-checking even though the work had been approved by the sample checks made by one of the flying squads. As a result of this instruction, a fair amount of approved work was double checked. In order to evaluate the efficiency of the sampling plan, we made a detailed investigation of the errors located in checking the approved work and contrasted them to those located in checking the rejected work.

This analysis proved conclusively that the sampling procedure had really succeeded in separating the inefficient teams from the efficient teams. In almost every case the double-checking of approved teams revealed only a bare minimum of errors, quite consistent with the AOQL used. The work of the rejected teams produced additional errors in over 90% of the cases examined. We found that, on the average, the approved teams counted about 30% <u>more</u> items in the same time period as the rejected teams, while committing <u>less</u> than one-fourth as many errors.

The store executives and auditors have been well pleased with this application of statistical sampling methods. The savings in direct labor costs were substantial (several thousand dollars annually). In addition, indirect savings resulted from the reduction in time required for the inventory.

In conclusion, I should add that the examples I have described represent only a few of the possibilities that exist in today's modern department store. Although our experiments have clearly demonstrated the potentialities of Statistical Quality Control, sample surveys, and other statistical procedures that I have not discussed, we have thusfar investigated only the more obvious applications. New techniques, as we all know, must advance slowly and soundly. We must gain the confidence of those involved, convincing them that the new procedure is superior to the old one. In the intensely competitive markets of today, only one answer really counts: Does the new procedure mean greater profits than the old one? This is the question that we asked at The Emporium. We are pleased with the answer.

ENGINEERING STANDARDS AND TOLERANCES

Ervin E. Schiesel
The Mattatuck Manufacturing Company

Solving the engineering of inspection problems in mass-produced parts or assemblies calls for a new concept in communications. Essentially, the blueprint with its tolerances and specifications is supposed to convey information from Engineering to Production, or in the case of purchased parts, from the consumer to the vendor. Because of the deficiencies of most engineering prints, Inspection or Quality Control interprets engineering data to Production and vice-versa, with its inspection procedures, quality level evaluations and test runs. Since blueprint specifications are incomplete, the details of manufacturing are being filled in by Inspection. This is a poor tribute to the engineering of products and has developed as a result of the highly-mathematical "theory of games."

Until the science of quality control uncovered the many secrets of machine capability, tool wear and process control, the determination of specifications and tolerances was the sole responsibility of engineering. In turn, the production of what appeared on the blueprint was the domain of the machine operator, tool-setter, lead-man and foreman in production. With only two parties concerned, the setting of tolerances was a comparatively simple game of designing to close limits with the expectation that these limits would be exceeded by ten to twenty per cent and still be suitable. If the limits were exceeded by more than this, the logical solution would not be to relax the limits for easier productivity, but to tighten them so that less parts would be acceptable under the new tolerances, thus tending to reduce the defects. The many ramifications of this simple game are well known and documented very well in many, many blueprint specifications and tolerances.

With the advent of control charts and sampling plans, the "game" of tolerance setting became more complicated in that quality control entered the scene with information that was not available to Production and Engineering, so that the strategy of setting realistic specifications was changed. This new information could be viewed and interpreted with considerable bias by either Engineering or Production, further strengthening their original concepts. In turn, quality control could team up with either side to make tolerances impossible or workable, depending upon the interpretation of results not known to either. At times, organization politics also served to complicate the use of supposedly impartial data, because the data takers may play a more active part in the game than is generally supposed. At best, this calls for grand strategy in a war that is too complicated and dangerous to play in industry. In order to eliminate the game aspects of specifications and tolerances, and in order to communicate better the requirements of the part for production with respect to fit, function and finish, it is recommended that in mass-produced parts and assemblies -

1. No dimension be placed on a blueprint unless it is to be measured or gaged by both Production and Inspection.

2. No dimension shall be placed on a blueprint <u>unless</u> the method of measurement or gaging is specified to all concerned.

3. No tolerance be placed on a dimension unless it can adequately

be measured or gaged for control.

4. No tolerances be placed on a dimension unless it can be supported by process capability studies.

5. No set of dimensions and tolerances is complete unless it spells out how the parts or assembly is to be used, and the overall check is if it is functionally correct for its use.

The above requirements appear basic, yet a cursory examination of most part prints will illustrate two or more deviations from the above list. The lack of good or uniform specification lies in the limited vocabulary of the blueprint language. These shortcomings, if corrected, can aid in again establishing Engineering as the real source of true and understandable specifications. In turn, Production and Inspection can produce, measure and evaluate realistically, rather than by interpreting the diplomatic-type language now used.

The basic communications are the use of gaging and measurement symbols developed by the Mattatuck Manufacturing Company. In addition, symbols for condition measurement are borrowed from the Ordnance Corps of the U.S. Army.

The following simple Greek letters and condition symbols connecting related dimensions now make the print a living document, which gives instructions and helps, rather than hinders understanding. The symbols as outlined are self-explanatory and are illustrated in the example of the bushing.

GAGING AND MEASURING SYMBOLS

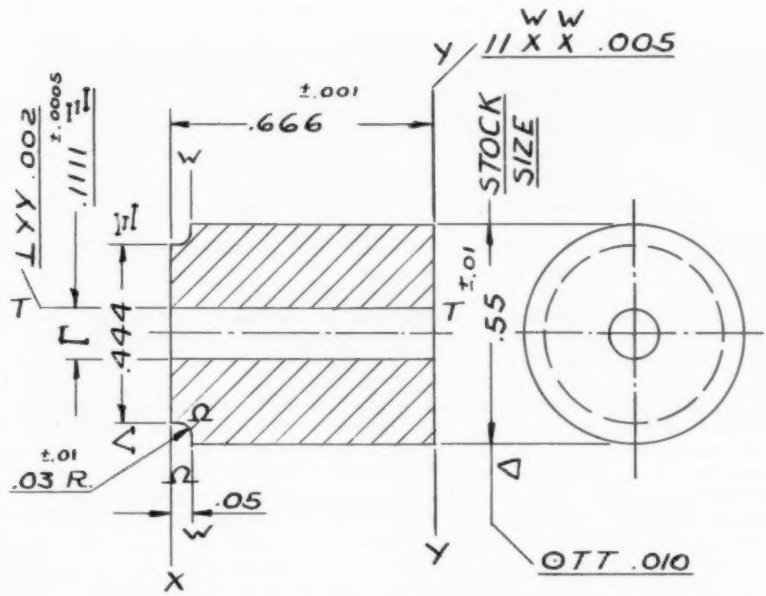
Δ	DELTA	Direct measurement by scale, vernier, micrometer, yardstick, etc., where actual measure can be made and discerned.
Γ	GAMMA	Go or No Go Gaging. Fixed gages such as plug, ring, thread or snap gages.
Λ	LAMBDA	Limit gages which indicate or compare dimensions between limits such as dial indicator or thread comparators.
Ω	OMEGA	Optical measurement or gaging by use of projection equipment and layouts where symbol is used. The tool must control the dimension.
Ξ	Xi	Combination gage which checks the two or more dimensions as used in an assembly or in relation to each other.

CONDITION MEASUREMENT SYMBOLS

//	PARALLEL	Symbols followed by one or more sets of reference letters indicate the lines, planes or surfaces to which the particular dimension
$\perp$	PERPENDICULAR	must be parallel, perpendicular or concentric. The figure following the reference letter indicates the tolerance for the characterist-
$\odot$	CONCENTRIC	ics involved.

THE MATTATUCK MANUFACTURING COMPANY
Waterbury, Connecticut

ILLUSTRATIVE EXAMPLE
BUSHING



Tolerance Guide:

1 Place Decimal	±.04
2 Place Decimal	±.003
3 Place Decimal	±.0002

Note:

Two dimensions to be checked with common gage using two go dimensions and no extra allowance for eccentricity. Part is used as gaged in assembly.

Tolerance recommendations are like cold cures; there are many, and none work better than the passage of time. However, these are the recommendations:

1. Use the number of digits in the dimension that are necessary in the tolerance.

2. Use bilateral tolerances in order that probability designs can be utilized more readily.

3. Use general guides to tolerances by the use of 1 and 2, in order to minimize the amount of tolerances shown after dimensions.

In order to change the present system, a whole new concept must be conveyed to Engineering. This, of necessity, will take time and should not be rushed by correcting old prints. New designs, however, should incorporate these symbols, even if only one at a time until Engineering and Production begin to understand the new communication means. This should aid in again establishing Engineering as the source of all information to Production and Inspection.

QUALITY CONTROL IN PAPER FINISHING

Edward R. Hoffman
Hammermill Paper Company

The Problem: To improve and control the perfection of manual counted and sorted paper.

The Process: So that the scope of the problem of defect removal and our approach to it may be appreciated, let us briefly review a few pictures of the process. Paper is made at Hammermill Paper Company complete from log to packed paper. Fig. I. views the process at the end of the paper machine, where a roll is being wound from the continuous web of paper coming from the machine. Here is a unit piece, one roll,

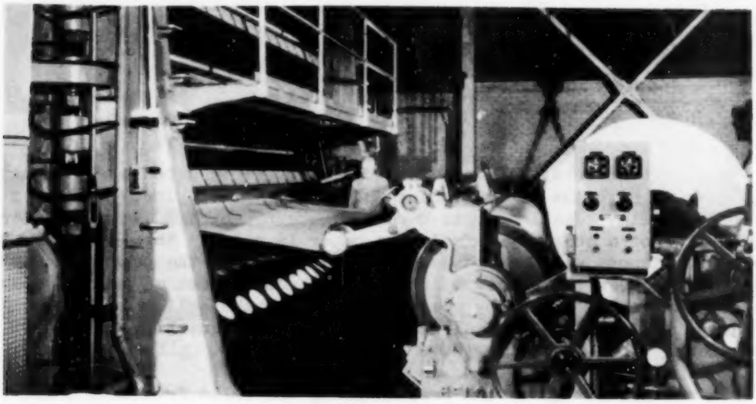


Fig. I. Paper Being Wound at Paper Machine

containing about 520,000 sq. ft. of paper or 8000#. Since the web from the machine flows out at about 800'/min., it must be stored or wound immediately into roll form. Defective paper as a result of starting a machine or changing grades can be disposed of by discarding parts of rolls formed during this period. Once the paper machine is adjusted and acceptable paper is being made, the operator attempts to prepare rolls of standard diameters. Uniformity of roll diameters is an important requirement in the efficiency and effectiveness of a following stage of manufacture, the sheeting process. Once a roll is partly formed defects that may now occur cannot be readily seen or removed at 800'/min. from the flow of the process. Following the papermaking stage, the rolls are rewound on a winder to cut them to final size and to remove defects occurring in quantity. This machine is shown in Fig. II.

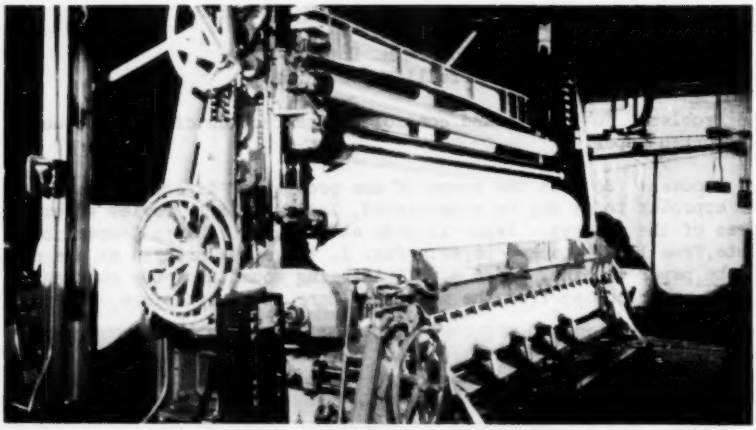


Fig. II. Rewinding a Roll of Paper

Although these winders can be stopped to remove defects, they must also keep up with the paper machine and generally do this by operating up to 3000'/min. Small defects hard to discern cannot be found. After the removal of a quantity defect, the paper is spliced together so as to form a continuous web for subsequent processing. In the paper industry, one must remember that a break in the paper web is just as much of a defect as a wrinkle. Thus, by removing a defect, the splice or sheet end is always left. It is logical then to remove only quantity defects at this point.

In the next stage of the process groups of rolls from 1 to 8 are sheeted at one time, Fig. III. Although these machines operate at a

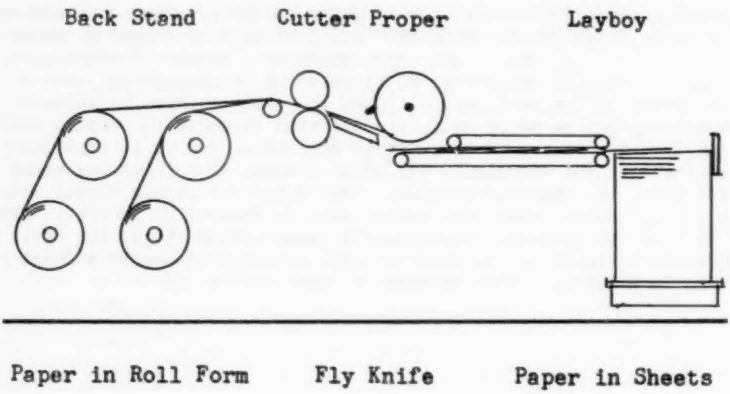


Fig. III. Schmatic Drawing of Sheeting Machine

moderate speed, the complexity of handling and watching several rolls at a time does not lend itself to good manual defect removal. As the

process is now operated, defective webs are interleaved with good paper.

Throughout rewinding and sheeting, defects are flagged with pressure glued paper markers. The subsequent cutting and trimming of the paper, however, cuts off the projecting end of a marker and/or separates a defect into several sheets, only one of which is marked. Flagging, although it helps in finding defects, is also a defect in itself if the flags are not found at a later point.

Shown in Fig. IV. is the next stage where the paper is manually

Fig. IV. The Counting and Fanning Area

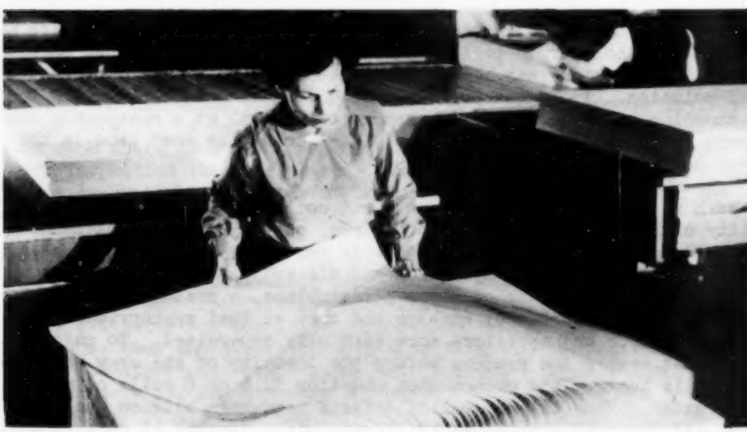


Fig. V. An Operator Fanning Paper

sorted for defects and counted. Fig. V. shows a detail of the fanning operation, and Fig. VI. the operation of counting. These girls experienced in the art of fanning paper and removing a sequence of defects,

Fig. VI. An Operator Counting Paper

say in every fifth sheet, do a remarkable job. Five hundred sheets or a ream is fanned and counted by an operator at the rate of one every 3.3 minutes. This figure includes the normal lost time typical of the operation. In this phase of the finishing operation there are 150 girls preparing 700,000# per day. The girls are grouped into 16 multiple unit crews each under the direction of a forelady. Two such units report to a first-line foreman. Two shift foremen and a superintendent heads the organization.

After counting and removal of defects, the paper is trimmed to final size, wrapped and labeled ready for shipment.

The Consideration of the Problem: To control quality in this process additional personnel for inspection purposes is required. It must be added wisely. Two questions can be asked to insure the eventuality of the proper solution. 1. Is the present method of manufacture conducive to good counting and good defect prevention or removal at a reasonable cost? 2. Where can inspection personnel be placed to do the most good in the present method?

With respect to method, let us first consider defects and the possibility of controlling this at the source, say by better maintenance and paper machine manipulation. There is certainly no profit in making a defective product in one operation and discarding it in the next. To obtain better care and use of the paper machines, a measure must first be made of the attributes by machine and crew so that appropriate repairs can be made and operators more carefully supervised. To gain this end one must sample the process before the identity of the crew who made the paper is lost. This occurs when sheeting 5, 6 or 8 rolls at a time on the cutter. Thus the removal of defects somewhere between the paper machine and the cutter would be the ideal area. Removal of defects at the paper machine or rewinders would be thwarted by the rapidity of the process and its continuous nature. Furthermore, removal of a defect at these points is impractical since this leaves a splice or sheet end which in itself is a defect. Also, defects occurring in the sheeting operation, would not be covered. Thus with respect to the identity of

the machine and crew, and to the easy removal of defects, the sheeting area is best. Here also speeds are considerably less and the process can be stopped if necessary. The only problem in this operation is to find and remove defects since the paper is already in motion. When we compare the efficiency of this possible method with the present method of hand fanning, requiring multiple units of girls, truckers, equipment and supervision, it is evident that development of a means of finding and removing defects at the cutter is highly desirable. The first recommendation, therefore, is to investigate and promote the possible equipment to remove defects at the cutter. This mechanism should also be arranged so as to count and mark each group of 500 sheets. Furthermore, it should provide the quality control group with a measure of the quantity of defects assignable to the various paper machine crews and machines. This is a long range program designed to improve the process in an economic fashion with the greatest benefits to the consumer and the company. Thus, the control of quality not only involves inspection, control charts and analysis, but it requires consideration of the process so as to make it more susceptible to quality control methods.

Since the process cannot be readily changed as described above, the second question is now ready for consideration. That is, where can inspection personnel be placed in the present process to do the most good? Since defects are removed and the paper counted in the fanning operation, a quality check must follow this operation and fall in the area before wrapping or before the identity of the operator preparing the paper is lost. This requires sampling immediately after trimming and at a point far enough from the operator so that it is difficult for an operator to anticipate an inspection. The longer the time or the number of units on a conveyor between an operator doing a repetitive manual job and the inspection area, the less is the inspection attention necessary to check an operator. This applies, of course, to cases where the operator's efforts are not self evident. The number of units on the conveyor between the two points can be counted and the time interval and inspection attention time calculated to keep the operator's risk of inspection as high as possible. With five units or blocks of paper on the conveyor system between the operator and the inspectors, an attention time of 16-2/3% is sufficient to keep the operator's risk of inspection at 100%. One inspector covering eight conveyors could devote 12-1/2% of her time to each. A decision was made, therefore, to place one inspector on each shift at a 75% operator's risk of inspection rather than put on more inspection than necessary. Once the inspector is assigned and the data is available, a study of the process can be made to get the best balance between the operation and inspection.

Data Handling and Reporting: In this operation of manual counting and sorting, generalized data would probably not control the quality to the desired degree and therefore requires that each individual girl be repeatedly evaluated for her ability to remove defects and to count. Once this is done then the supervisors of manufacturing could begin to control the operators' efforts. Since two inspectors could readily make 200 inspections per day covering 150 girls, 16 crews, etc., a procedure to analyze and record data had to be developed. A single entry card system was resorted to and a card developed that could be sorted and analyzed readily by hand. If an acceptable gain in quality could be demonstrated, then a commercial system of data analysis might be justified. Fig. VII. shows the card developed. On this card the shape of the

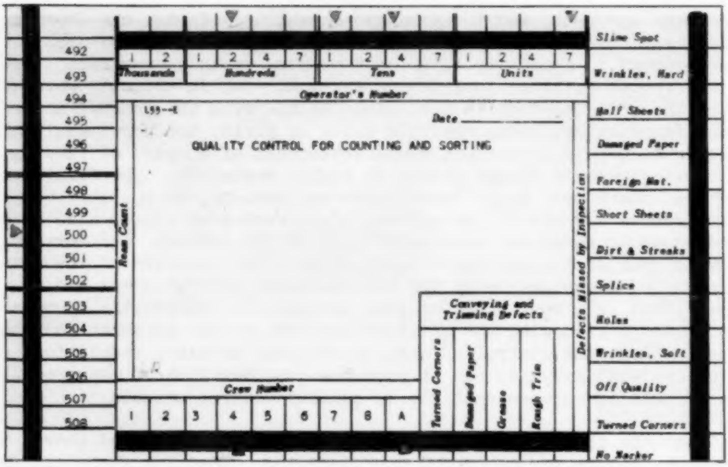


Fig. VII. Single Entry Data Card for Count & Defects

marginal punch indicates the inspector. Fig. VIII. shows a group of cards displaced for rapid analysis.

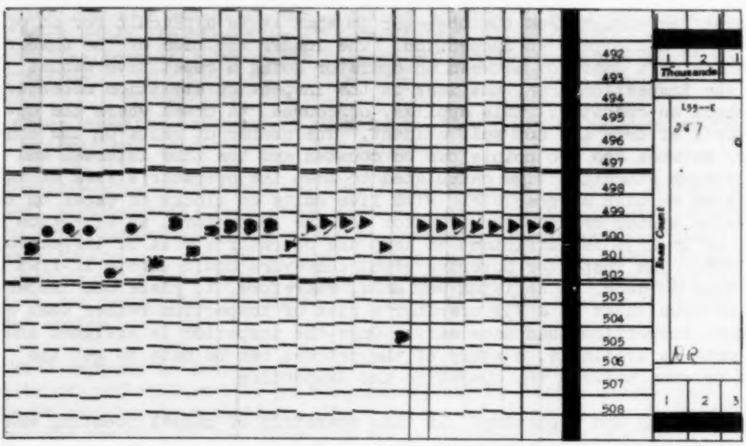


Fig. VIII. Displaced Cards Ready for Analysis

These cards can be rapidly sorted by hand for inspector, for operator by number, by crew and shift. A group of cards may be fanned out as in Fig. VIII. and the number of counts at the various values determined. The possibility of rapidly determining the per cent miscount, the distribution of count errors or per cent defects is obvious. When large groups of cards are to be counted, an electronic counter of the stylus type is used.

Although the 5 x 8 cards may be conveniently accumulated by groups in an ordinary file, it is essential to keep the cards moving through

the analysis system so that data is frequently available in all phases. Furthermore, all the raw data had to be made available to the operators and foremen because they were interested and concerned about the records of their efforts. This was accomplished by transferring the data from the cards to a loose-leaf notebook sheet using a form as shown in Fig. IX. One or more sheets as required for each operator are arranged in the notebook in operator number order. At the end of each shift the inspector arranges the inspection cards for the shift in number order by the punch system and transfers the data. Next, the range is calculated for count and the % defective for defects. The group size is 20 with each group overlapping the previous group by 10 observations. Corresponding 6 sigmas are written in for each group by use of a precalculated table.

Since it is necessary to keep the raw data posted in the Finishing Room, these records are also used to make 3 month summaries by operator. The summary contains average % defective, average count range and % miscount based on a specification of 500 $\pm$ 2. The records of the 150 individuals are kept on 5 x 8 file cards with data listed for each.

Summaries by operators made every three months have been found to be a satisfactory rate at which to accumulate the information. Generally speaking, the values do not fluctuate too much and the trends toward good or poor performance are gradual. One can now very well depend on an operator not to pass more than her established level of defects, nor to exceed her established count range. One other thing is done to keep the individual operator fully informed. That is, "A Notice of Excessive Defects or Miscounts" is issued to anyone exceeding the normal distribution of all operators for a group size of 20 consecutive inspections. On these notices, the total number of notices sent to the individual and the number of consecutive notices prior to the present notice are recorded. This helps supervision to temper their discussions with the individuals.

This covers in detail the analysis of data with respect to the individual operators. Concurrent with this the single entry data cards are moved on through the analysis system.

In the next phase of analysis the cards are accumulated by conveyor in groups of 100 cards. From the groups a calculation and a report of per cent defective and per cent miscount are made and plotted on control charts. Results are reported at the rate of 2 conveyor groups per day. General summary control charts are also issued by shift and General summary control charts are also issued by shift and for the entire operation after each round of charts for all conveyors. All charts are prepared by the Quality Control Group and mailed to the supervising group. On the back of each chart, spaces are provided for issuing date, signature of supervisor and comments. The charts are also returned by mail. With each chart a bar-graph is included giving in detail the breakdown of the defects passed by that group. For purposes of general supervision, control charts are plotted for the various defects based on the entire operation. This gives them the opportunity to know if any one item needs special attention along the line of new instructions or new equipment.

With respect to the data card again, all cards are finally sorted into operator order and filed in boxes by the month. Since each card is signed by the operator at the time of inspection, these serve to verify

Operator No. 5

Date		NOV 10 1953	NOV 11 1953	NOV 17 1953	NOV 23 1953	NOV 25 1953	DEC 2 1953	DEC 8 1953	DEC 16 1953
Counting									
492								I	
493									
494									
495			I						
496									
497									
498			I						
499									
500	𝍸 III	𝍸 𝍸	𝍸 II	𝍸 𝍸	𝍸 III	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸
501					I				
502									
503	I				I				
504						I			
505			I						
506	I								
507									
508									
R	6	0	10	0	3	4	0	8	0
6σ	11.7	5.9	9.8	9.8	2.9	6.8	3.9	7.8	7.8
Defects Missed by Inspection									
Slime Spots									
Wrinkles, Hard		I							
Half Sheets									
Damaged Paper									
Foreign Material									
Short Sheets									
Dirt & Streaks									
Splices									
Holes									
Wrinkles, Soft									
Off Quality									
Turned Corners			II						
No Marker									
No. of Defective Beams		I	II						
No. of Inspections	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸	𝍸 𝍸
% of Beams with Broke	0	5	15	10	0	8	0	0	0
		NOV 6 1953	NOV 10 1953	NOV 17 1953	NOV 23 1953	NOV 27 1953	DEC 2 1953	DEC 8 1953	DEC 16 1953

Fig. IX. Form for Operator Analysis

the inspection and our calculations. Cards are held for 3 months before discarding.

For the purpose of determining the process average count, range and standard deviation, a sample of 400 cards is drawn at random from the cards accumulated for each month. Through the punch system, the number of counts in each count class is readily determined and from this the distribution of the process. In the case studied the average number of sheets per ream decreased from 501.8 to 500. The high initial count may be contributed to excessive positive errors and related to the fact that a customer is seldom critical of this kind of defect. The savings in paper has been sufficiently great so as to more than justify the cost of inspection and analysis.

Results: The program accomplished the objective, that is, the improvement of counting and defect removal and has maintained this for a long enough period so as to establish control. Graphically the results may be shown in Fig. X. and Fig. XI. Fig. X. shows the drastic improvement in count and the large per cent now falling within $\pm$ 2 counts. Fig. XI. shows a like improvement with respect to defects with near perfection for periods as long as 10 days frequently occurring.

Besides accomplishing the above improvement, the program also resulted in the delivery of packages more nearly containing 500 sheets with a savings in paper that more than supported the program.

Conclusion: In the application of statistical quality control to an established process, a study of the process is important to determine how it might be changed so as to be susceptible to statistical methods. Too frequently statistical methods are distorted to fit the process. This approach will at least assure a sound long range program.

For the present process it is possible to predetermine inspection attention time so as to keep the operator's risk of inspection at a high level. Such a determination is useful in deciding the minimum required inspection. With minimum inspection, ample analysis made possible through a rapid sorting, single entry punch card system and supervision's cooperation, it is possible to control with assurity a large scale manual operation.

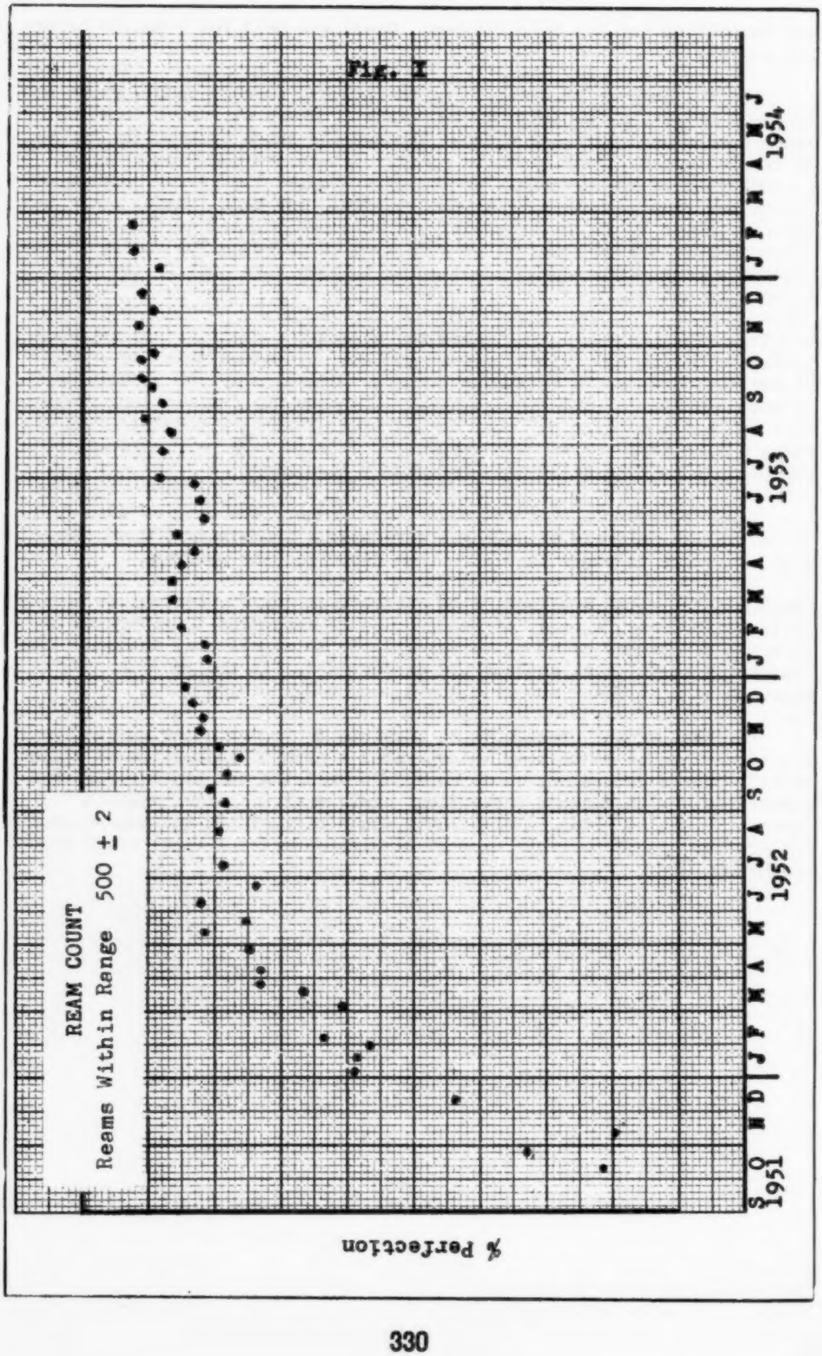


Fig. I
REAM COUNT
Reams Within Range 500 ± 2
% Perfection
S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J
1951
1952
1953
1954

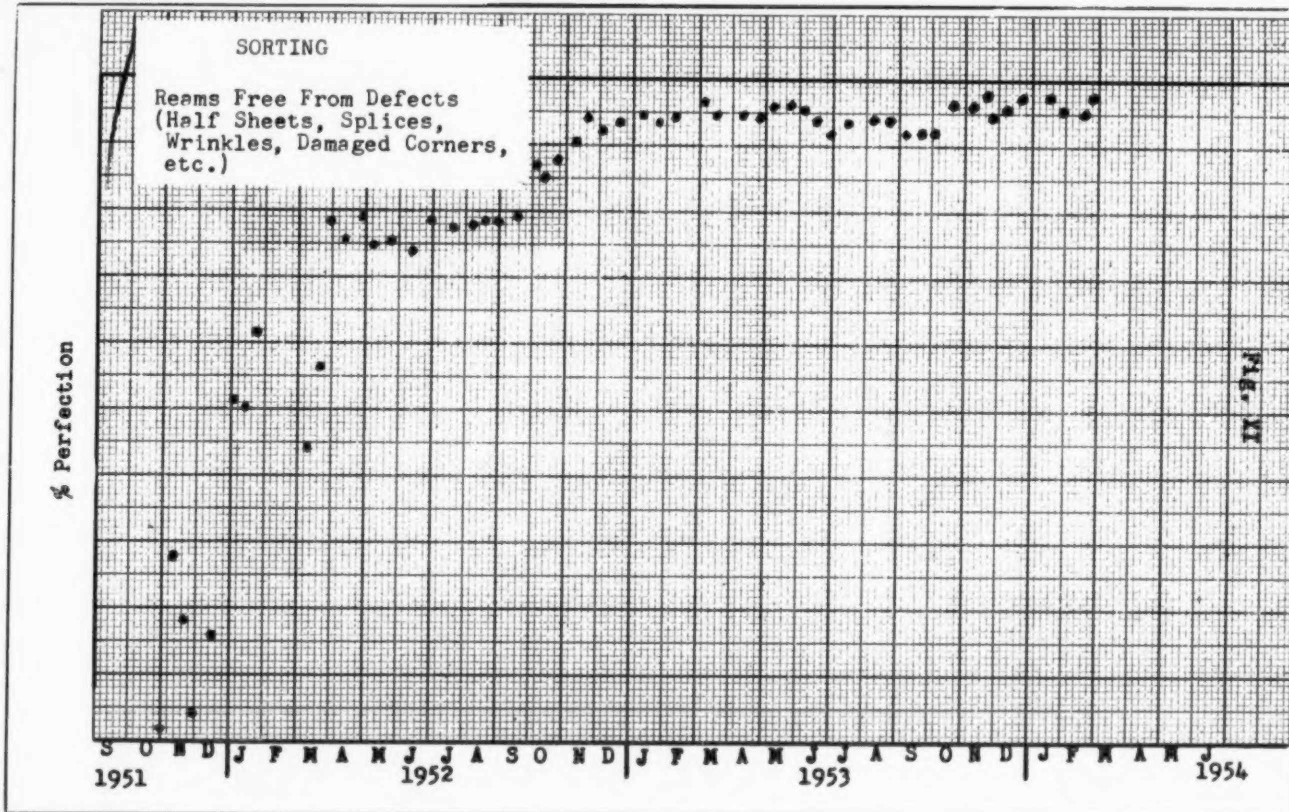


Fig. XI
SORTING
Reams Free From Defects
(Half Sheets, Splices,
Wrinkles, Damaged Corners,
etc.)
% Perfection
S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J
1951
1952
1953
1954

S.Q.C. APPLICATIONS FOR IMPROVING AIRCRAFT AND ENGINE MAINTENANCE

W. W. Wilcox and A. M. Hull
United Air Lines

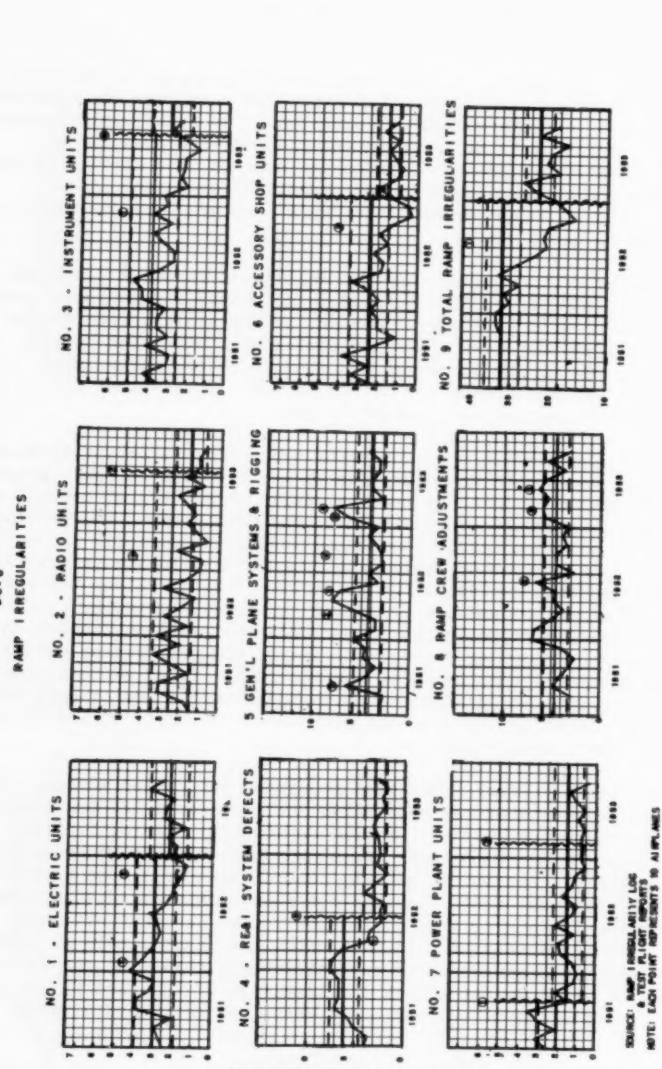
Any transportation company, an air line in particular, which does its own maintenance and overhaul work is its own customer from the standpoint of the quality of the overhauled product turned out of its maintenance bases. Units and systems of the airplane checked or overhauled at United Air Lines' San Francisco Maintenance Base go into service on our own airplanes and are monitored by our own flight crews and Line service mechanics. These men are "customers" who are very exacting in their demands on our product.

The quality of the product that we deliver to you as a passenger flying United Air Lines or as a shipper is of course a function of many variables, all of which are very important to the satisfaction of the exacting "Air Age" customer. This paper will deal with that portion of the quality of our product which has to do with the mechanical dependibility, air worthiness and safety of the airplane that is the final means of delivering our product to you.

A progressive airline has a complete system for reporting in detail the mechanical aspects of the operation of the airplane to some central base of operations. Flight log reports, daily telemeters of mechanical deficiencies of importance to Engineering and Maintenance, and records of findings of the line service mechanics during the periodic inspections are the principle sources of basic data from which valuable information may be summarized and evaluated for corrective action. Since the mid 1930's United Air Lines has had organized mechanical irregularity summaries on its various types of airplanes and about 1943 we started on a program of making irregular replacement charts and performance curves on the many removable components of our aircraft. Today we have some 1500 of these graphs and histograms which indicate the performance of the principle operating units of the airplanes. Mr. Hull will tell you more about these charts later.

In 1947 shortly after I had been appointed Inspection Manager of our San Francisco Maintenance Base I had the good sense, if you please, to enroll in the University of Iowa Ten Day Quality Control Training Program offered by our well known Dr. Lloyd Knowler. Our progress in the application of statistical quality control methods to the work of our shops and overhaul activities has not been spectacular but as you will see in the few examples we shall show you we have met with some real successes. It can safely be said that with this tight circle of information and details coming back to the maintenance base from our field operations we certainly don't lack for knowledge of the general and the specific problem areas of our various types of aircraft and aircraft equipment.

It might be claimed that we don't need a set of control lines to tell us we're in trouble with a particular unit or category of irregularity. There's no comfort in the knowledge that you have a performance process in a beautiful state of "statistical control" but at much too high a level!



MANAGEMENT S.Q.C. CHART
FIGURE 1

Out irregularity charts for years were used without benefit of control limit applications and were very valuable. However, Mr. Hull will tell you how we have sharpened that tool by the use of SQC. Therefore, while we do not lack for information on our trouble areas, we still have the problem of defising fixes for our deficiencies either through design improvement, improved overhaul procedures, training of personnel, materials improvements, and changes of operating specifications to more realistic levels, these problems open up many specific applications of quality control methods. Problems like evaluating the specific performance and calibration of a unit, the success of a proposed design change by a service test, the evaluation and detailed study of critical quality characteristecs from several different sources of supply - all of these things call for intelligent use of the statistical methods that we all are learning about through the fine efforts of this Society. I shall now proceed to show you some of the applications we have made of SQC within our Maintenance Base organization. Mr. Hull will show you our applications in the field of engineering analysis, particularly as it applies to the performance of our aircraft engines and accessories out in line service.

FLIGHT LINE QUALITY CONTROL

One excellent measure of the quality of our efforts is to log, summarize and portray the findings of the flight line crew and test pilots. When an airplane leaves one of our overhaul docks a separate crew, called a ramp crew, takes over. When this crew completes a varying amount of unfinished dock work they then begin their primary function of checking the airplane, engines, and all the systems in preparation for test flight. We then have an operation that produces quality control information on the collective efforts of all the inside shops and the overhaul dock. Deficiencies discovered and corrected by the ramp crew plus those reported by the test pilots are summarized by airplane type and become a series of "C" charts.

Figure No. 1 shows a typical management chart that is a "C" chart with each point representing the average number of defects for ten airplanes. For example the graph portraying radio unit removals includes all those units removed from the airplane for cause which are from the radio overhaul shop under a particular supervisor. The two system defects charts (#4 and #5) are a measure of the quality of the overhaul dock work since system troubles come from defects within the airplane itself.

Back of this slower moving management "C" chart is a whole family of C charts where the unit is only one airplane. These charts record defects per airplane on ramp and test flight and detailed foot notes of assignable causes are kept. Thus, we know the chronic troubles that are occurring on the ramp operation and on test flight.

[illegible]

[illegible]
[illegible]
W. E. Wilcox

x
x x
x

6554 7-18-52

RE&I Ixxxxx Ixxxxxxxxxxxxxxxxxxx

[illegible]

MD [illegible] cabin press. limit control - Polarization of receptical off

Found - insert had Rotated slightly in shell - Replaced Receptacle

[illegible] "Dock Removal" stamp!

MD. [illegible]-123 - [illegible] alt. fuel quantity indicator - Too sensitive [illegible]. change under vibration.

The jewel mounting on pointer bracket was loose.
Since this is a manufactured assembly, it was necessary to replace the pointer bracket assy.

MD. 4485-12 compression ratio limit switch - Blowing circuit breaker

The unit came back completely [illegible]. Opened case and found [illegible] wire burned in half. Have contacted Eng'g and standardized internal wiring on the unit and all future 4485's [illegible]

MD 4344-65 supercharger oil temp. ind. - Right indicator reads too high

Insulation of "Right" Armature Windings "Broke Down".
This is another case where we must rely on the quality of material the Lewis Engineering Co. uses in manufacturing these units, since we have no way of knowing when this "Break Down" condition is approaching.

G.M. [illegible]

MD 4069-43 Pedestal, controller (turn) - elevator trim [illegible] meter sticks when brought full down

Report was correct. Shop test procedure for this unit has not, up to now, included full scale deflection of this 3-axis meter. We are now making this test and have set up [illegible] for testing this meter the same as if it was removed from the controller.

MD 4273-160 outside air temp. indicator - Erratic oscillates several degrees

This unit was given a complete shop check + it tested OK.

MD 7288 cowl flap relay - Shorting or grounding (Power Contacts)

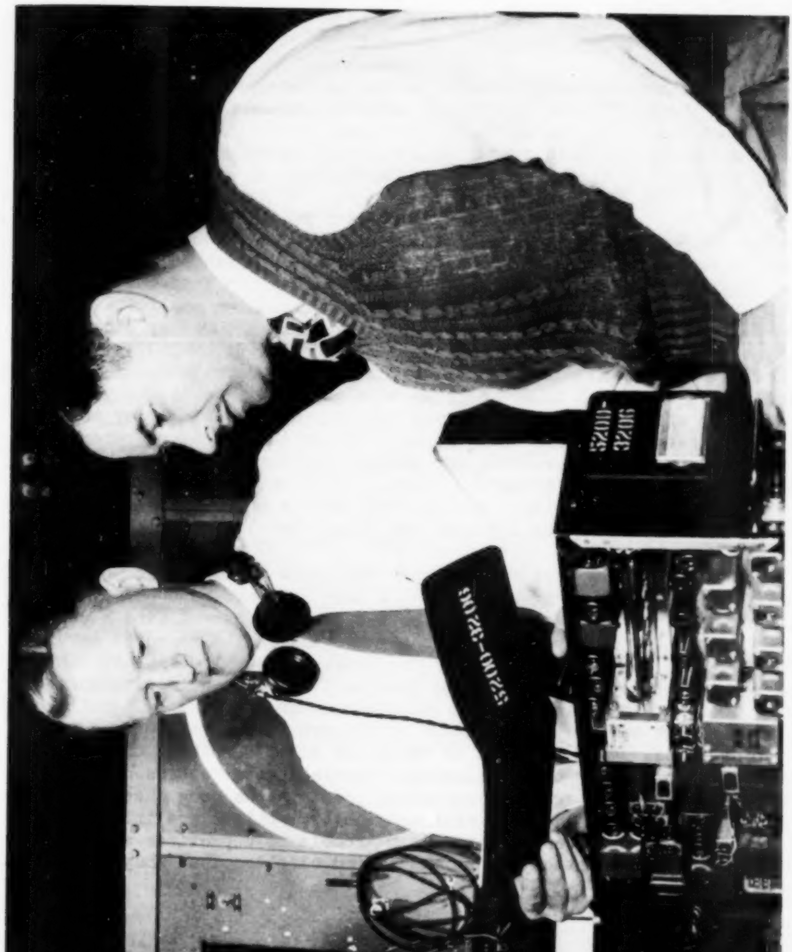
[illegible] G.M. [illegible]

Power Contact ceramic [illegible] were carboned and one was cracked. This is very common on all [illegible] removals of the ADI relay. This relay had never been in the shop.
(over)

CHALLENGE REPORT
FIGURE 2

But charts without follow-up action will not improve a process. An out of control point on a detail "C" chart causes a Challenge Report (Fig. 2) to go to the shop responsible for the type of unit covered by the chart. Since the reproduction of the introductory statement on the report is so poor I shall quote It:

"A high number of "RE&I" shop units (or system) defects have been reported by the Ramp and Test Flight crews while the above airplane was in for overhaul. It appears to be worth investigation into



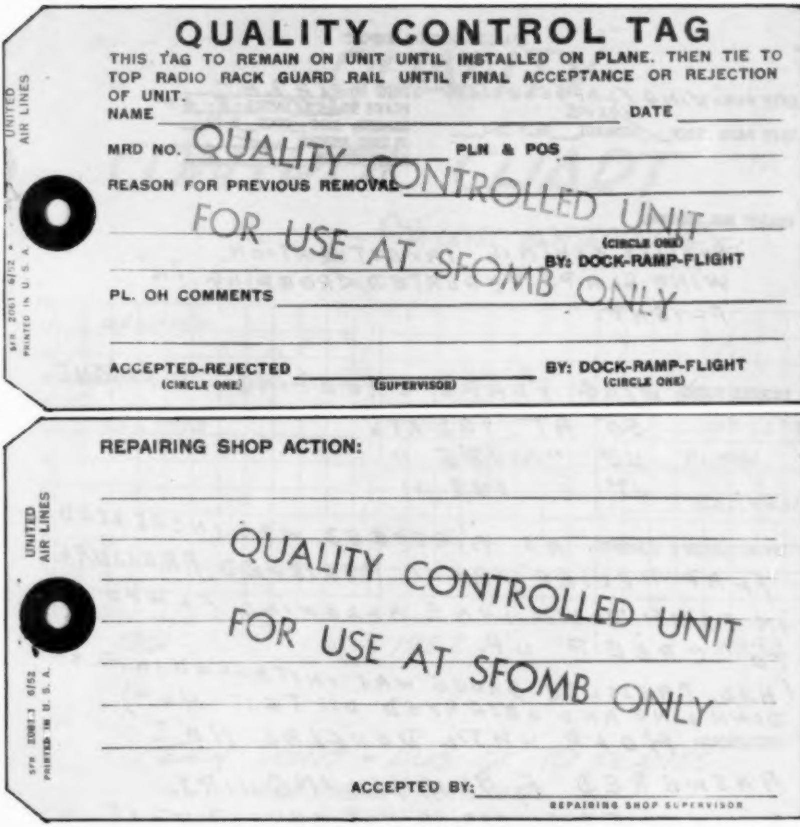
UNITED'S FIRST QUALITY CONTROL TEAM - 1951
FIGURE 3

the specific units or system reported to determine whether there is a correctable condition which is in your power and responsibility to correct. Please summarize your recommendations or corrective action taken below each item and return to Inspection and Quality Control Division thru your manager."

These challenge reports serve to advise the overhaul shop of the nature of the operational deficiency on each unit removed. The shop knows that the ramp removed "zero" time unit is coming back to the shop. The shop supervisor then gets into the touble shooting work on these units and reports his actions to his foreman and to Quality Control.

This challenge report system has worked reasonably well but we really hit pay dirt when we thought of the idea of "a quality control team". This team consists of two men - one an experienced man from one of the inside shops; the other an experienced man from one of the dock crews or the ramp crew, or the San Francisco line station. These men serve a period of about a month and then are relieved by other men and work on a temporary assignment under my assistant, Mr. R. R. (Bob) Nichols. They spend full time following through on units removed from airplanes by our ramp crew. Working with the shop experts, they coordinate shop efforts with the ramp and test flight crews in order to really ferret out basic causes of our troubles and make firm recommendations to shop supervision and Engineering for effective corrective steps.. The efforts of these Q.C. Teams and of the shop supervisors and engineers working with them has done more to uncover the real reasons for our weak spots in quality than any other phase of our program.

Another wrinkle that we have invented which may interest you is the Quality Control Tag. (Figure 4) When a unit has been removed any time within the first ten hours of service and the shop cannot find a cause for malfunctioning they attach this tag as well as the standard serviceable tag when the unit is returned to stock. This white Quality Control Tag has the effect of quarantining that unit to be held in Maintenance Base stock for installation only on airplanes in the Base. It will not be shipped out to our line stations as a spare unless it is for a plane out of service. When the unit is installed on an airplane, the installing mechanic returns the regular serviceable tag to records but mounts the quality control tag at a convenient point in the crew quarters of the airplane. Thus, the ramp crew and the test pilots are informed of any of these units which have given trouble on a previous airplane and need particular attention on the second airplane that they have been installed on.



UNITED AIR LINES

SFR 2061 6/52 PRINTED IN U. S. A.

QUALITY CONTROL TAG

THIS TAG TO REMAIN ON UNIT UNTIL INSTALLED ON PLANE. THEN TIE TO TOP RADIO RACK GUARD RAIL UNTIL FINAL ACCEPTANCE OR REJECTION OF UNIT.

NAME ______ DATE ______

MRD NO. ______ PLN & POS ______

REASON FOR PREVIOUS REMOVAL ______

(CIRCLE ONE)
BY: DOCK-RAMP-FLIGHT

PL. OH COMMENTS ______

ACCEPTED-REJECTED (CIRCLE ONE) ______ (SUPERVISOR) BY: DOCK-RAMP-FLIGHT (CIRCLE ONE)

QUALITY CONTROLLED UNIT
FOR USE AT SFOMB ONLY

UNITED AIR LINES

SFR 2061.1 6/52 PRINTED IN U. S. A.

REPAIRING SHOP ACTION:

QUALITY CONTROLLED UNIT
FOR USE AT SFOMB ONLY

ACCEPTED BY: ______ REPAIRING SHOP SUPERVISOR

QUALITY CONTROL TAG
FIGURE 4

QUALITY CONTROL REPORT

UNIT NAME: WING FLAP SELECTOR VALVE

UNIT BASE CHECK ✓ OVERHAUL___ UNIT TSO___

PART NO: 8522-

PLANE NO 657 DATE 2-X-54

REMOVED: RAMP ✓ DOCK B-29___

FLIGHT REPORT ✓

RAMP REPORT___

REASON ~~FOR REMOVAL~~:

QUALITY CONTROL INVESTIGATION. WING FLAPS REPORTED CREEPING IN FLIGHT.

TROUBLE FOUND: WING FLAPS CREEPING IN FLIGHT.

50° AT 105 KTS
45° " 135 "
43° " 148 "

INVESTIGATOR'S COMMENTS: AS AIRSPEED WAS INCREASED FLAP RELIEF VALVE RELIEVED PRESSURE IN FLAP DOWN LINE ALLOWING FLAPS TO CREEP UP.
(HYD. PRESSURE GAUGE WAS INSTALLED IN FLAP DOWN LINE AND OBSERVED ON TEST HOP.)

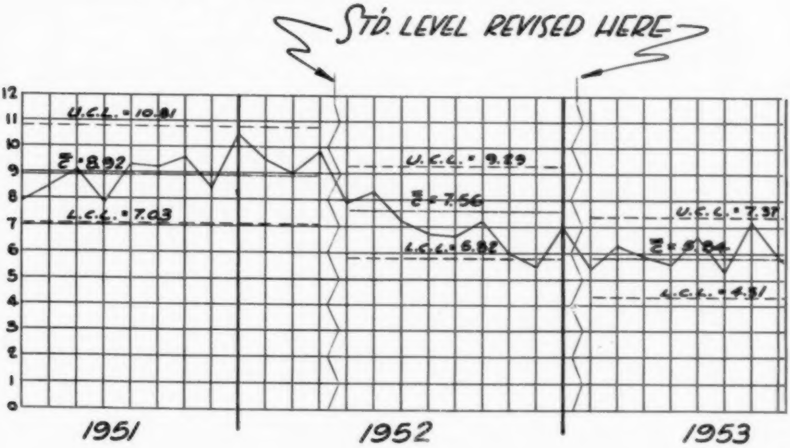
CONCLUSION: HOLD UNTIL DOUGLAS HAS ANSWERED F. BLUMS INQUIRY.
FLIGHT TEST ENGINEERING HAS AGREED THAT THIS IS A NORMAL CONDITION AND WILL NOT GRIPE DC-6B'S IN THE FUTURE. DACO CONFIRMED THAT THIS IS NORMAL ON DC6-B

INVESTIGATOR'S NAME Spornaugle
DATE 2-XI-54

FIGURE 4A

Figure 4A is a typical quality control team report. We have books full of these. In this case a test flight "squawk" of the tendency of landing flaps to creep up slightly during landing approach speeds was carefully studied by the Quality Control team. Working with the ramp crew and the hydraulic systems engineer, and after consultation with the Douglas Aircraft Company, the Q.C. Team persuaded the Flight Test Manager that this was a normal condition designed into the flap hydraulic system to avoid excessive air loads on the flap structure. Another expensive "defect" was resolved objectively and another reason to re-fly the airplane was eliminated.

QUALITY CONTROL CHART



EACH POINT = AVG. OF 10 PLANES

SHOP FOREMAN'S Q. C. CHART
FIGURE 5A

SHOP APPLICATIONS OF SQC

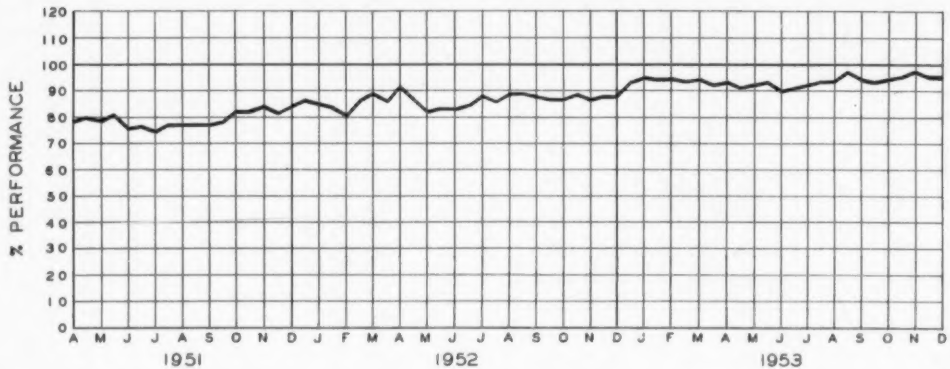
The Radio Electrical and Instrument Shop has many active Q.C. charts

Each Lead Mechanic keeps a P chart on the out-put quality of his crew. Figure 5A shows a C chart for Ramp Irregularities of all units overhauled in the RE&I shop. The quality level by mid 1953 indicated a need for another lowering of C and this improvement has been maintained into 1954.

Figure 5B is a chart of this shop's labor performance or efficiency rating. These charts show that quality of out put can be steadily improved at the same time that other management efforts are successfully improving the out put efficiency.

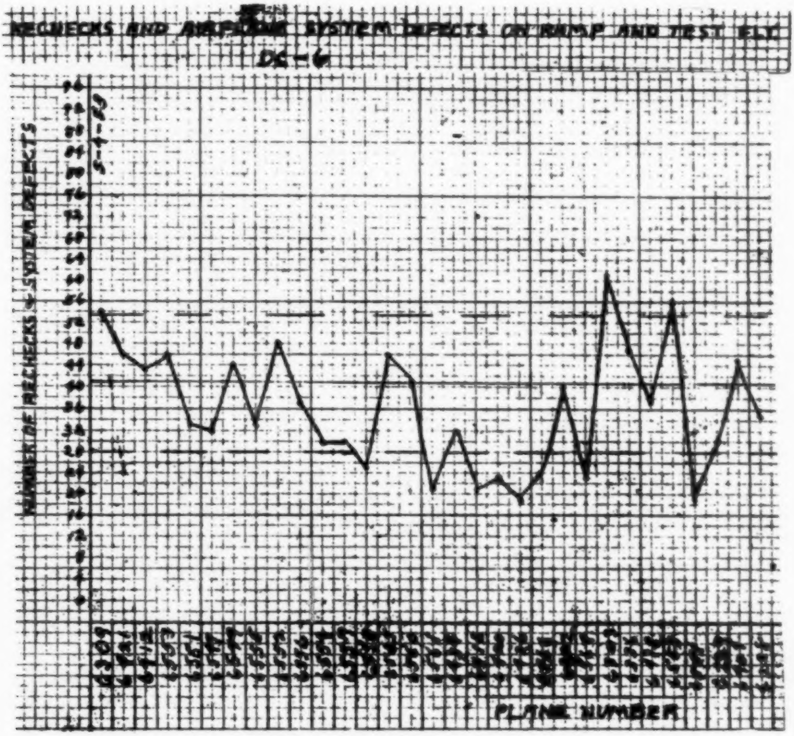
R.E. & I.

LABOR PERFORMANCE



THE LEVEL OF PERFORMANCE AGAINST WORK STANDARDS REFLECTS SHOP SUPERVISION, WORKER OUTPUT, MATERIAL FLOW, AND OTHER MANAGEMENT CONTROLLABLE VARIABLES.

FIGURE 5B

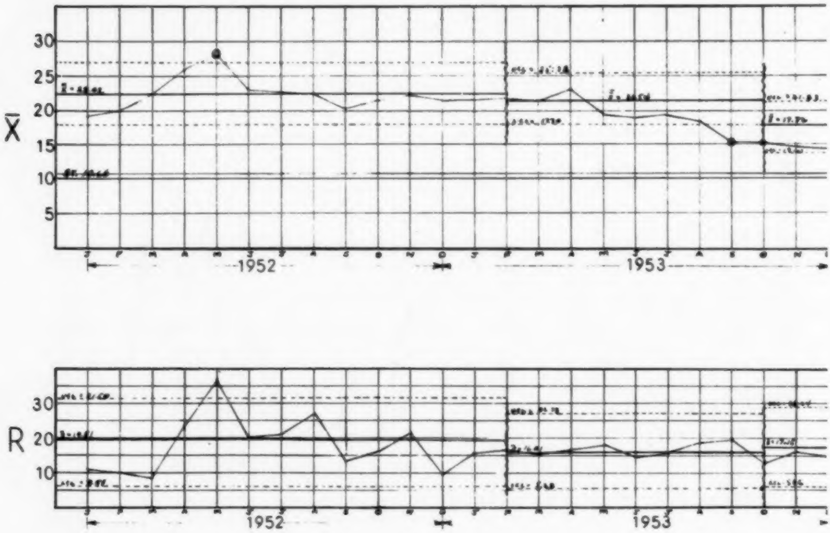


TYPICAL DOCK "RECHECK" CHART
FIGURE 6

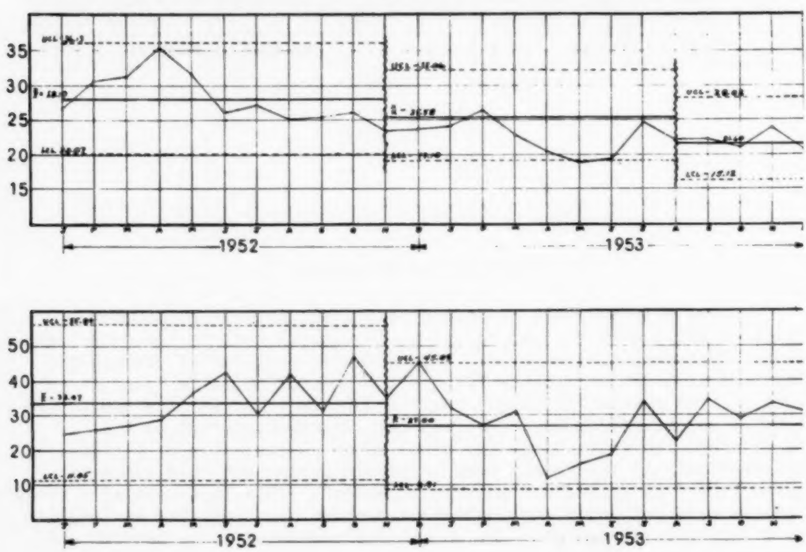
OVERHAUL DOCK OPERATION

Figure 6 illustrates a typical shop chart showing the number of rechecks written by our inspectors on mechanical work in the dock operation. Follow-up action on this type of chart consists of posting the detailed list of rechecks on the Quality Control bulletin board at each dock so that the mechanics and their supervisors can see the repetitive type of errors being made. We also keep "recheck charts" on the inspectors. By definition, any write-up by an inspector of a defect in an area that had previously been inspected during our "preliminary inspection " is called a "recheck-inspector". This is a sensitive subject but has made mechanics and inspectors recheck conscious and we feel it is a factor in encouraging our mechanical personnel to be careful and thorough in their work.

DC-6 Routine Airplane Inspection Manhours



DC-6 Routine Engine Inspection Manhours



$\overline{X}$ + R MANHOUR CHARTS
FIGURE 7

Our San Francisco line station has assigned one of our SQC students as a staff man to the Line Maintenance Manager. He maintains quite a series of charts, some of which use control limits to trigger management action. Figure 7 shows two X and Range charts recording the "Routine" man hours spent on airplane inspections and engine inspections respectively. "Routine" man hours are those required to do the periodic inspection of the airplane and the standard work items such as a scheduled spark plug change. It does <u>not</u> include the "non-routine" repair, adjust, or replacement work. The downward trend results from intensive training and a close Work Standards program.

Last year incoming Convair 340 flights to San Francisco that were scheduled to have only a light "Terminating Check" actually were "loaded" with discrepancies. First by "C" charting these flight reports (Figure 8) for the facts then showing headquarters our problem we applied pressure for various types of corrective action with a steady improvement to the February '54 level. Each point on the graph, incidentally, is the total "squaks" on 5 airplanes.

TERMINATING PREFLIGHT DISCREPANCIES

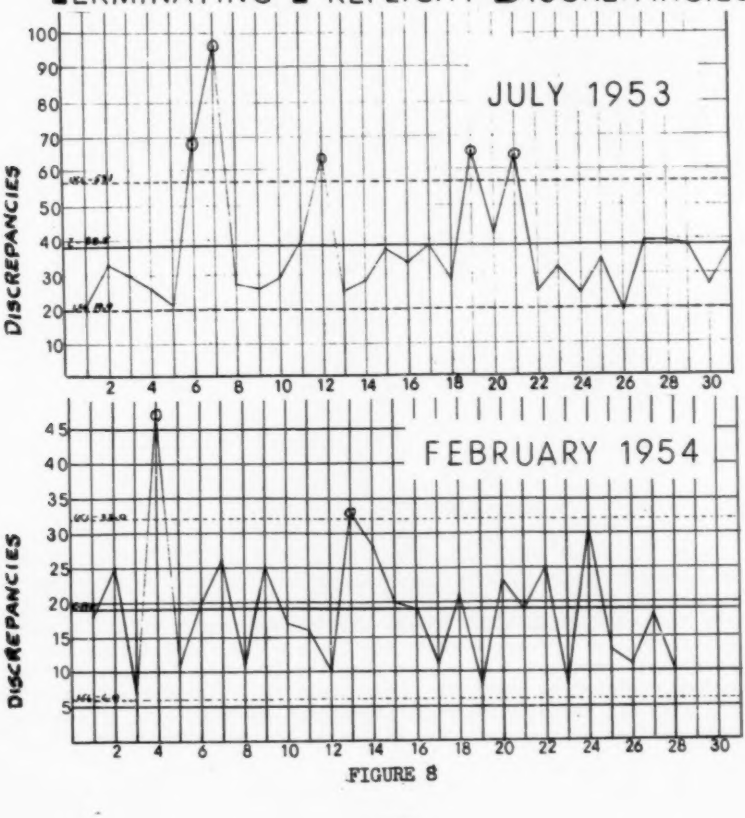


FIGURE 8

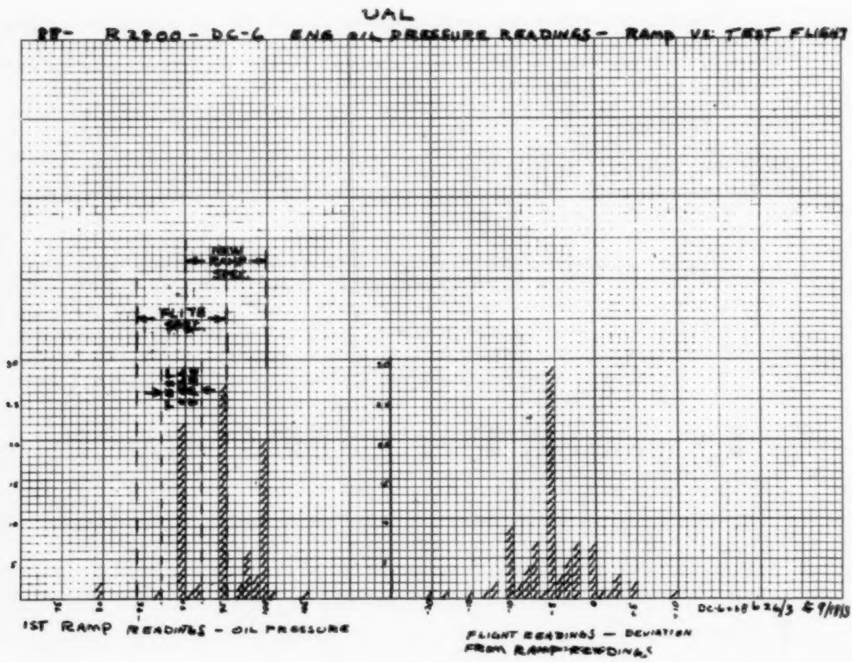
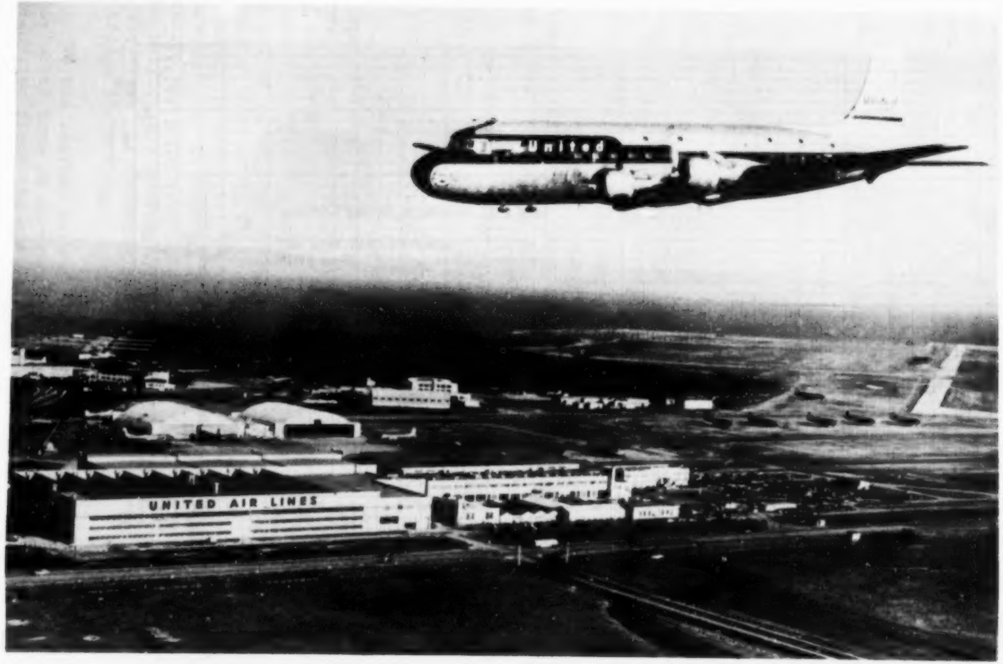


FIGURE 10

DETAIL STUDIES OF TROUBLES

Figure No. 10 is an illustration of one of many simple histograms that we have made on particular variables in our adjustment processes on the settings of airplanes' and engines' instrumentation. To explain, note the vertical dotted lines at the left of the page showing the span of the test cell oil pressure specification and the wider flight and ramp wider specification. Note also that the pressure readings as taken by the ramp centered at about the upper limit of the ramp specification. The ramp crew was making a downward adjustment on about half of the engines after the initial ground run. The histogram to the right shows the pattern



BIRD'S EYE VIEW OF UNITED'S SAN FRANCISCO MAINTENANCE BASE

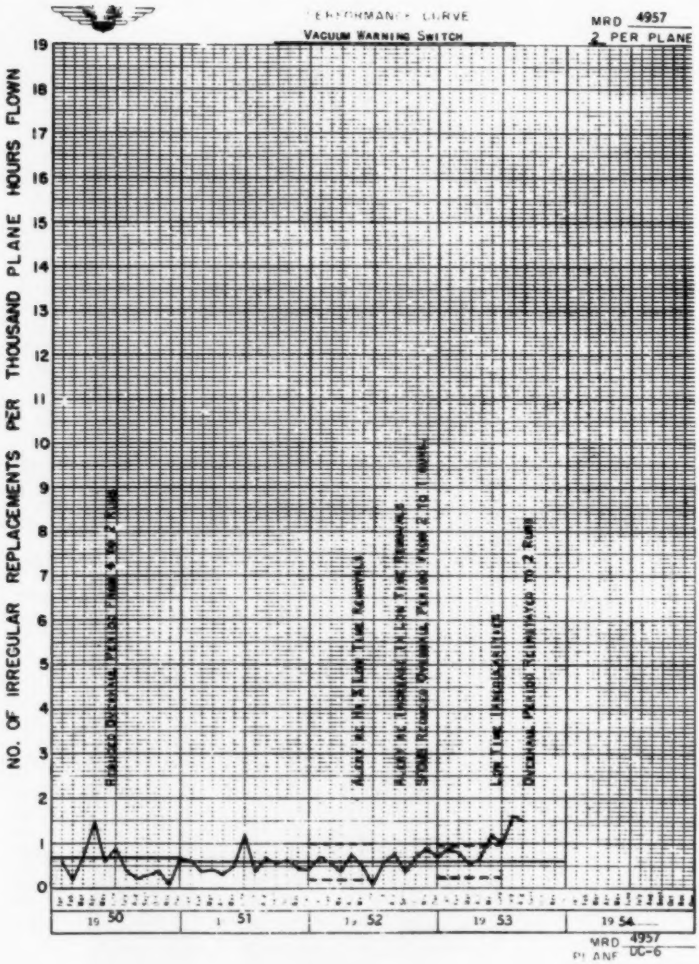

PERFORMANCE CURVE
VACUUM WARNING SWITCH
MRD 4957
2 PER PLANE
NO. OF IRREGULAR REPLACEMENTS PER THOUSAND PLANE HOURS FLOWN
19 50
19 53
19 54
MRD 4957
PLANE DC-6

FIGURE 11

of deviation of the test flight reading from this final ramp crew setting on the ground. For some reason the test pilots were reporting an oil pressure reading (at a given oil temperature) averaging 5 lbs. lower than the reading of the ramp crew. Result was that the ramp crew's adjustment downward of the engines out of the engine the test cells were being squawked by the pilots too often and they often had to re-adjust the oil pressure upward. The ramp crew is now instructed to center their oil pressure adjustments at 95 ± 5 lbs. on our confident expectation that most of these readings will be read lower by the test pilot and within the official specification of 90 ± 5 lbs.

Other histograms of the many variables in our business have been made with real success in getting to the heart of a problem and correcting it.

TRAINING PROGRAM IN S.Q.C.

We have emphasized training of shop people and engineers in basic principles of S.Q.C. rather than trying to build up a large quality control group. I can say now that I can no longer keep up to the new applications of quality control devices that are going on in production shops and at our line station at SFO. About 100 people have been trained, or are being trained, within the Base and now in the Stores and Accounting Departments.

This training approach to spreading the use of SQC methods may be slower than other methods but it's a "grass roots" approach to the problem and therefore is more accepted by the men on the job.

Mr. Hull will now explain some of the application that he has made with S.Q.C. in his Engineering Analysis work.

AIRCRAFT AND ENGINE PERFORMANCE CURVES

One of the earliest measures of aircraft performance was the Performance Curve. This curve expresses performance in terms of "Number of Irregular Replacements per Thousand Plane Hours Flown." In the course of the past 10 years, this curve has become the aircraft industry's "yardstick" for measuring or comparing performance. Until recently, we relied on individual judgment as to whether or not the monthly fluctuations were of sufficient magnitude to warrant investigation; it was a matter of personal opinion, tempered by the experience of the individual interpreting the curves, as to whether or not the curve was out-of-control and warranted investigation. The addition of control limits to the performance curves (see broken lines Figure 11) has eliminated the "guess work" in determining whether or not the variation is significant - a triggering action.

We had learned through experience that the monthly fluctuations of the curves bore an inverse relationship to the fleet size - i.e., the larger the fleet the smaller the fluctuation; and conversely, the smaller the fleet the greater the fluctuation - clearly obeying the laws of probability. We interpreted our problem as a "defects per unit" problem, in which the number of irregular removals was interpreted as the number of defects, and a thousand hours flown as the unit. In other words, the number of irregular removals per thousand hours flown is analagous to the number of defects per unit. The unit, instead of being an airplane or an airplane overhaul period, was visualized as a thousand hours of exposure -

CONTROL LIMITS
FOR
UAL PERFORMANCE CURVES

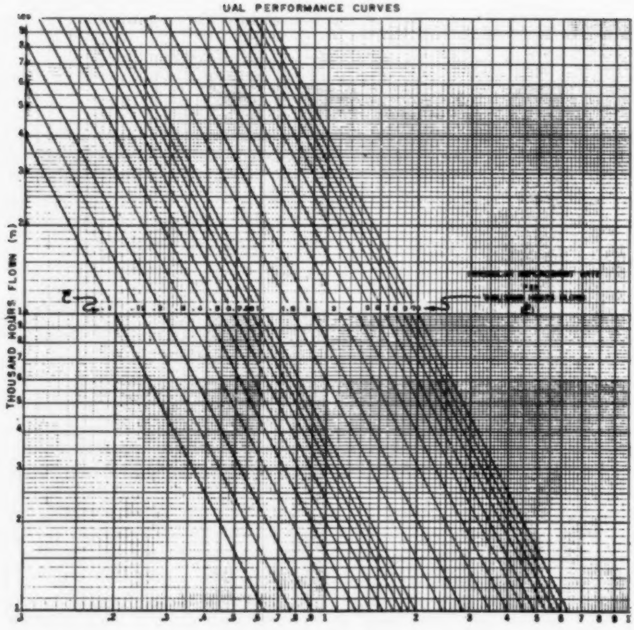


FIGURE 12

the denominator we had been using for years. Since our problem reflected the influence of fleet size, the need for an "n" in the formula was evident. Therefore, the formula $\bar{\bar{C}} \pm 2\sqrt{\bar{\bar{C}}/n}$ was used, in which $\bar{\bar{C}}$ (C double bar) is the previous year's average irregular removal rate, and n, the thousand of hours flown per month.

NOMOGRAPH FOR CALCULATING CONTROL LIMITS

Since we maintained performance curves on approximately 1500 units, roughly 300 units for each fleet, conventional methods of calculating control limits were too cumbersome and time consuming. To facilitate calculations, we drew the nomograph covered by Figure 12, based on the formula $\bar{\bar{C}} \pm 2\sqrt{\bar{\bar{C}}/n}$, in which $\bar{\bar{C}}$ (C double bar) is the previous year's average Irregular Replacement Rate per Thousand Hours Flown, and n, the Thousand Hours Flown per month by the particular fleet of airplanes. For a component with an average irregular replacement rate of 0.6 (six tenths) per thousand hours flown, the control limits are calculated as follows (reading from Figure 12):

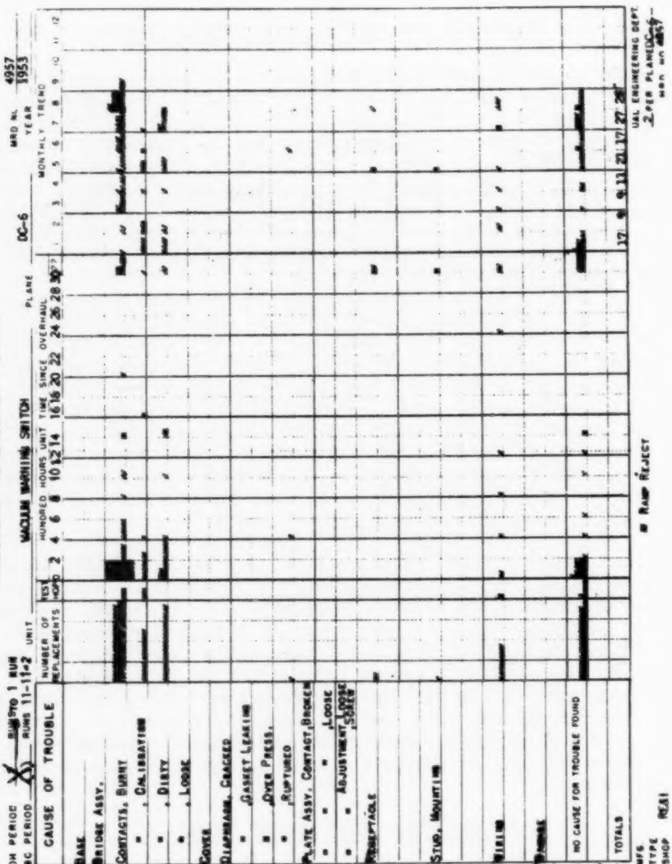
For our DC-6 fleet, which averaged 15,000 hours per month during 1952, follow the line n = 15 Thousand Hours Flown until it intersects the line C = 0.6. Read the ± Control Limits from the abscisse where the two points intersect. This value is read ± 0.4; therefore, the control limits are 0.6 ± 0.4, or an upper control limit (UCL) of 1.0 and a lower control limit (LCL) of 0.2 (see Figure 1 for 1952).
For our B-377 fleet, which averages 1000 hours per month, follow the line n = 1 Thousand Hours Flown until it intersects the line C = 0.6. The control limits are ± 1.55. The upper control limit (UCL) is 0.6 + 1.55, or 2.15, and the lower control limit (LCL) is 0.6 - 1.55, or zero in this case.

Since we wanted our control limits to give us advanced warning of impending trouble, we used two sigma (2σ) control limits. For the advanced warning, we felt we could afford to be wrong one time in twenty, or 5% of the time. However, since we had information readily available as to "Casue of Trouble" for each irregularity (see Irregular Replacement Chart), we were in a favorable position to judge whether or not there was an assignable cause for the out-of-control condition. Furthermore, an out-of-control condition for two consecutive months at 2σ control limits gave a slightly higher probability of an assignable cause than one month out-of-control at 3σ control limits. In addition, we had one month's advance warning.

IRREGULAR REPLACEMENT CHART

We developed the Irregular Replacement Chart (see Figure 13) as a quality control tool approximately eight years ago. With minor modifications, it has been adopted by several domestic and foreign airlines, by aircraft and component manufactureres, as well as by a number of industries. The green "Repairable" (shop repair) tags covering irregular replacements (premature or unscheduled removals) are posted as follows:

Cause of Trouble - Each irregular removal is analyzed and classified as to the primary cause of the difficulty or irregularity. The trouble is posted by making a slash (/) mark under each of the following three headings, opposite the appropriate Cause of Trouble.
Number of Replacements - This column develops into a histogram that quickly shows the number of irregular replacements for each cause. Entries are made left to right starting at the lower left hand corner.



IRREGULAR REPLACEMENT CHART
FIGURE 13

Hundred Hours Unit Time Since Overhaul - This column shows where in the overhaul period the irregularities are occurring, localizing the problem for expedient and effective corrective action. Irregularities that occur in the early portion of the overhaul period are suggestive of a shop, installation, or handling problem; whereas irregularities occurring late in the overhaul period indicate that the overhaul period is subject to question (see discussion on Overhaul Period Analysis Curve).

Monthly Trend - This column shows the month-to-month trend for each "Cause of Trouble." This section of the chart localizes the difficulty calendarwise, gives an indication of seasonal problems, and gives an advanced indication of impending chronic trouble.

Frequently, trouble areas are localized and corrective action effected before the Performance Curve has shown an out-of-control condition - an out-of-control condition may be avoided altogether when expedient action is taken. The Irregular Replacement Chart gives a complete, but condensed, graphic picture of component operating problems so that the cause of trouble can be readily isolated for effective and expedient corrective action.

To keep our shop and engineering personnel informed of out-of-control conditions, and of troubles warranting corrective action, we forward them a print of the Irregular Replacement Chart and the Performance Curve, along with a brief explanation. When this file is returned with shop and engineering comments, it is reviewed, further action initiated as warranted, and filed in a "Chronic Unit File". Subsequent action at monthly performance reviews is based on the performance trend of the unit, and the comments previously noted in the Chronic Unit File. The history file on the unit is routed along with each recurrent report to give the complete history on the unit. In addition we send copies of the Irregular Replacement Chart and Performance Curve to the component manufacturer to keep him advised of service problems involving his components.

OVERHAUL PERIOD ANALYSIS CURVE

We developed the Overhaul Period Analysis Curve as a tool for determining the optimum overhaul period of aircraft components. This curve expresses performance on a mortality or actuarial basis, using Thousand Hours Flown per Irregular Removal as the base or measure of performance (see ordinate of Figs. 14 and 15). The Overhaul Period Analysis Curve answers two questions commonly asked when analyzing the performance of a component. These questions are: (1) What is the performance in each TSO Bracket (age group or cell) of the overhaul period? and (2) When should the component be overhauled for optimum performance?

Care must be exercised in selecting the sample for the analysis, first, to insure that it is representative of the problem, product, or process under consideration; and, second, that the sample is of sufficient size to insure a representative cross-section of the product. In general, we have found that a calendar period equivalent to the overhaul period of the component, represented by at least 100 removals, time and unscheduled combined, constitutes the minimum sample size. Transitional periods should be avoided, periods during which there was a modification, or a change in process or specifications. In such cases it is desirable to take a sample both before and after the change, and to compare the results. This gives a valid measure of the two processes and provides a scientific basis for management decisions.

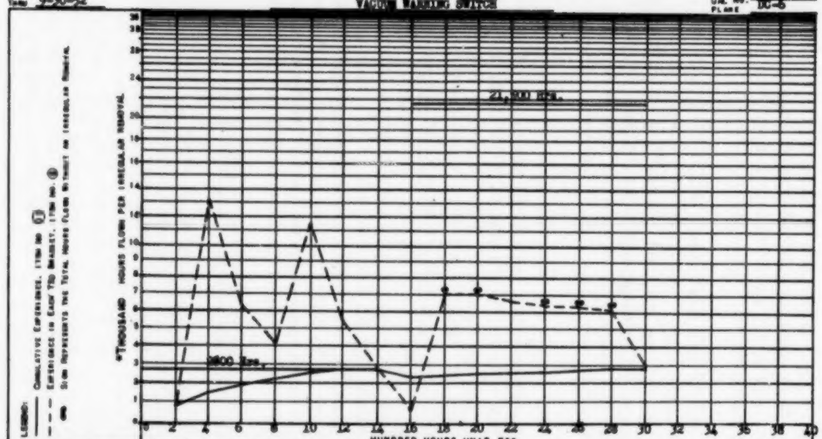
OVERHAUL PERIOD ANALYSIS CURVE

FOR

VACUUM WARNING SWITCH

FROM 1-1-52
THRU 9-30-52

UAL NO. 4957
PLANE DC-6



	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	TOTAL
1 NO. IRREGULAR REMOVALS	#) 36	#) 2	4	#) 6	2	#) 4	#) 7	#) 19			1				#) 1						82
2 NO. SCHEDULED REMOVALS	2			3	2	1	11	36			2		1	1	29						88
3 NO. LEFT IN SERVICE	132	130	126	117	113	108	90	35	35	35	32	32	31	30	0						170
4 100 HRS. ACCUM. ON 1	36	2	4	6	2	4	7	19			1				1						
5 100 HRS. ACCUM. ON 2	2			3	2	1	11	36			2		1	1	29						F
6 100 HRS. ACCUM. ON 3	264	260	252	234	226	216	180	70	70	70	64	64	62	60	0						
7 Σ 4+5+6	302	262	256	243	230	221	198	125	70	70	67	64	63	61	30						48%
8 7 ÷ 1 = EXP. IN BRACKET	8.4	131	64	41	115	55	28	6.5	70	70	67	64	63	61	30						
9 ITEM 7 CUMULATIVE	302	564	820	1063	1293	1514	1712	1837	1907	1977	2044	2106	2171	2232	2262						
10 ITEM 1 CUMULATIVE	36	38	42	48	50	54	61	80	80	80	81	81	81	81	82						
11 9 ÷ 10 = CUM. EXPERIENCE	8.4	14.8	19.5	22.1	25.9	28	28	22.9	23.8	24.7	25.2	26	26.8	27.6	27.6						

* NOTE: TELESCOPIC SCALE (ARITHMETICAL PROGRESSION)

#) = Ramp Removals

MAINTENANCE ANALYSIS SECTION
ENGINEERING DEPARTMENT
UNITED AIR LINES
JCL 11-18-52

OVERHAUL PERIOD ANALYSIS CURVE
FOR VACUUM WARNING SWITCH
FIGURE 14

The <u>non-cumulative curve</u> (see broken line Figs. 14 and 15) shows the mortality rate in each TSO bracket (age group or cell) of the overhaul period. The TSO brackets are represented on the abscissa (horizontal scale) of the curve in Hundred Hours Unit TSO (Time Since Overhaul). Performance for the cell interval is plotted on the right hand margin of the cell. This non-cumulative curve shows the TSO brackets in which performance is unusually good or bad. It focuses attention to a particular portion of the overhaul period. By referring to the Irregular Replacement Chart, Figure 13, we can evaluate the problem in terms of the nature and severity of the irregularities in each age group. When the non-cumulative curve stays above the cumulative (solid line) curve discussed below, the performance is better than the average performance to that point. When the performance in the individual age groups falls consistently below the cumulative curve, the non-cumulative curve (see non-cumulative curve Fig. 15 above 5000 hours unit TSO) indicates that performance in these age groups is poor, has deteriorated, and is below the average to that point. This curve indicates that the overhaul period is too long, and for optimum performance should be reduced. The experience in each TSO bracket is found by dividing the total hours operated by all units (the exposure) in the TSO bracket by the number of unscheduled changes (the experience) in the TSO bracket.

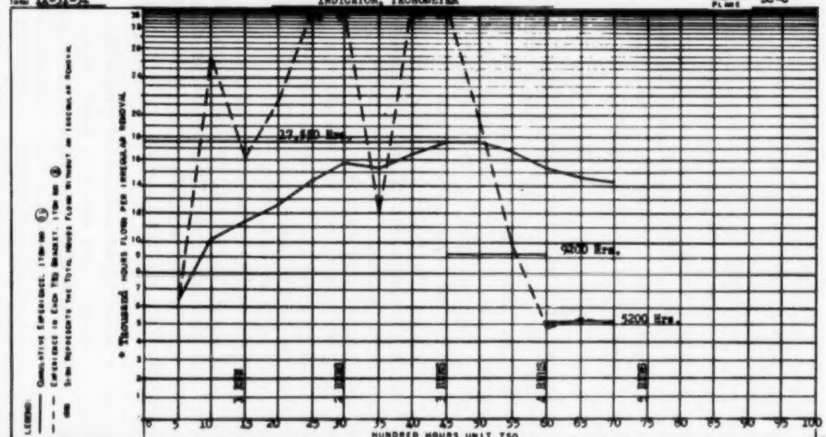
The <u>cumulative curve</u> (solid line) shows the cumulative experience at any given point in the overhaul period. It shows the performance of all units from zero time (overhaul or new)to each point in the overhaul period represented on the abscissa. These points also are plotted to the right hand margin of the cell. Performance at each point in the overhaul period is calculated by dividing the total hours flown by all units to that point in the overhaul period by the total irregular removals during the same period. The result is the hours flown per irregular removal to that point in the overhaul period. In other words, if the sample taken is a representative sample, and represents the process or product being produced, each point on this curve represents the performance that can be expected if the unit is overhauled at that particular point in the overhaul period. For optimum performance, the component must be overhauled at the highest point on the cumulative curve. If this point occurs at the end of the overhaul period, a further increase in overhaul time may be warranted. A positive slope to the curve may be an indication that the overhaul period should be extended, while a negative slope may indicate that the overhaul period should be reduced. The final decision is based on the safety and economic factors involved.

Figures 14 and 15 represent a simplified method of preparing an Overhaul Period Analysis Curve, a method that we have been using very successfully for several years. The Irregular Removals are tallied according to the Unit TSO bracket or cell in which the irregularity occurred under item 1 , <u>Number of Irregular Removals</u>. Likewise, the Scheduled or Time Removals are tallied according to the cell representing the TSO bracket in which they were removed, under item 2 , <u>Number of Scheduled Removals</u>. <u>The Number Left in Service</u> at the end of the TSO bracket or cell, item 3 , is found by subtracting the number of irregular and scheduled removals (item 1 plus 2) from the total units in service at the beginning of the TSO period (this is the "Number Left in Service", item 3 , at the end of the previous TSO period). Item 4 is found by multiplying item 1 by one-half ($\frac{1}{2}$) the cell interval in hundred hours - this method simplifies carrying decimals. One-half the cell interval is used since it is assumed for simplicity that, on the average, the removal took place at the cell mid-point. For example, if

OVERHAUL PERIOD ANALYSIS CURVE
FOR
INDICATOR, TACHOMETER

From 1-1-52
Thru 12-31-52

UAL No. 4413-4423
Plane DC-6



		5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	Total
①	No. Irregular Removals	18	4	6	4	2	2	6	1	1	3	5	8	5	2							67
②	No. Scheduled Removals	5	3	27	6	1	13	3	8	8	9	18	14	21	38							174
③	No. Left in Service	218	211	178	168	165	150	141	132	123	111	88	66	40	0							241
④	100 Hrs. Accum. on ①	45	10	15	10	5	5	15	2.5	2.5	7.5	12.5	20	12.5	5							
⑤	100 Hrs. Accum. on ②	12.5	7.5	67.5	15	2.5	32.5	7.5	20	20	22.5	45	35	52.5	95							[illegible]
⑥	100 Hrs. Accum. on ③	1090	1055	890	840	825	750	705	660	615	555	440	330	200	0							
⑦	Σ ④+⑤+⑥	1147.5	1072.5	973	865	833	787	727	683	638	585	498	385	265	100							
⑧	⑦ ÷ ① = Exp. in Bracket	63.8	268.1	162.2	216	416	393.6	121.1	683	638	195	99.5	48.1	53	50							
⑨	Item ⑦ Cumulative	1148	2220	3193	4058	4890	5677	6405	7087	7725	8310	8808	9193	9458	9558							
⑩	Item ① Cumulative	18	22	28	32	34	36	42	43	44	47	52	60	65	67							
⑪	⑨ ÷ ⑩ = Cum. Experience	63.8	100.9	114	126.8	143.8	158	153	165	176	176.8	169	153	146	143							

NOTE: Telescopic Scale (Arithmetical Progression)

Maintenance Analysis Section
Engineering Department
United Air Lines
FWJ 2-15-3

OVERHAUL PERIOD ANALYSIS CURVE
FOR
TACHOMETER INDICATOR
FIGURE 15

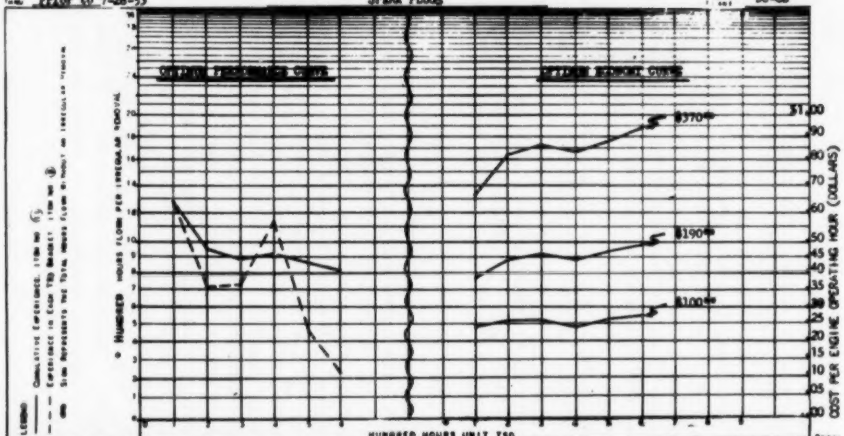
the cell interval is one (1) Hundred Hours, item 1 is multiplied by ½ or .5; if the cell interval is 2, item 1 is multiplied by 1; if the cell interval is 3, item 1 is multiplied by 1.5, etc. Item 5 is calculated by multiplying item 2 by ½ the cell interval, as in item 4 . Item 6 is calculated by multiplying item 3 by the cell interval in Hundred Hours since these components remained in service for the full cell interval. Item 7 is the sum of items 4 plus 5 plus 6. This figure represents the Hundreds of Hours operated by all units in that cell or TSO bracket. The mortality rate of each cell (TSO bracket or age group) is found by dividing the total hours operated by all units in the cell (item 7) by the number if Irregular Removals (casualities, item 1) in that cell, expressed as Hundreds of Hours. The result is the Experience in the TSO Bracket (item 8), it is non-cumulative, and is plotted as a broken line curve (- - -). When there are no irregular removals in a particular TSO bracket, the total hours operated in the bracket are posted to the curve with an infinity sign (∞) over the point. This symbol indicates the number of hours operated without an irregularity. Item 9 is the cumulative hours flown to that point in the overhaul period. It is found by adding the previous cell totals in item 7 to that point in the overhaul period. Item 10 is the total irregular removals to that point in the overhaul period. It is the cumulative totals in Item 1 , left to right, to that point in the overhaul period. Item 11 is the Cumulative Experience by all units to that point in the overhaul period - - the performance we would expect if this were the overhaul period of the component. Item 11 is calculated by dividing item 9 by item 10, and plotting it as the solid line curve, see Figures 14 and 15.

In addition to helping Management arrive at the optimum overhaul period, the Overhaul Period Analysis Curve serves many other useful purposes. For example, from the data section of the Overhaul Period Analysis Curve, it is possible to calculate the performance for any TSO increment of the overhaul period by dividing the total hours flown during the interval (found by adding item 7 for the interval) by the total irregular removals during the interval (found by adding item 1 for the interval). In this manner it is possible to compare the performance of any portion of the overhaul period with any other portion of the overhaul period. Also, the percentage of units remaining in service at any point in the overhaul period is found by dividing the Number Left in Service, item 3 , by the total number of units in the sample. Likewise, we can determine the average unit TSO of the units in the sample by dividing the right hand figure in item 9 , the Total Hours Cumulated by all units, by the total units in the sample. The percent defective ($\bar{p}$) for the sample;i.e. the percentage of irregular removals, is shown in the right hand margin of the data section.

Another adaptation of the Overhaul Period Analysis Curve is its use in determining the most economical overhaul period, for plotting an "Optimum Economy Curve". In airline operations, the cost of a delay is a very nebulous figure, one that can invoke considerable controversy. However, by assigning an arbitrary, an average, or a known cost figure to each irregularity, and "Optimum Economy Curve" can be drawn. Figure 16 shows one of these curves for spark plugs based on a somewhat arbitrary figure representing the cost of each spark plug irregularity. The lower curve is based on an average cost of \$100 per spark plug irregularity. This represents a one-half hour out-of-service revenue loss, \$94, plus \$6.00 manpower cost of replacing and overhauling the spark plugs. The middle (\$190) curve represents one hour out-of-service revenue loss, and

OVERHAUL PERIOD ANALYSIS CURVE
FOR
SPARK PLUGS

From Last Plug Check
Thru Prior to 7-28-53



								Total
①	No. Irregular Removals	22	30	19	7	9	9	96
②	No. Scheduled Removals	46	49	46	47	14	21	223
③	No. Left in Service	251	172	107	53	30	0	319
④	100 Hrs. Accum. on ①	11	15	9.5	3.5	4.5	4.5	P
⑤	100 Hrs. Accum. on ②	23	24.5	23	23.5	7	15.8	30.1%
⑥	100 Hrs. Accum. on ③	251	172	107	53	30	0	
⑦	Σ ④+⑤+⑥	285	211.5	139.5	80	41.5	20.3	
⑧	⑦ ÷ ① = Exp. in Bracket	12.9	7.1	7.3	11.4	4.6	2.3	
⑨	Item ⑦ Cumulative	285	496.5	636	716	757.5	777.8	
⑩	Item ① Cumulative	22	52	71	78	87	96	
⑪	⑨ ÷ ⑩ = Cum. Experience	12.9	9.5	8.9	9.2	8.7	8.1	

* NOTE: Telescopic Scale (Arithmetical Progression)
** Cost per Irregularity

Maintenance Analysis Section
Engineering Department
United Air Lines

OPTIMUM PERFORMANCE AND OPTIMUM ECONOMY CURVE
FOR
SPARK PLUGS
FIGURE 16

the upper ($370) curve a two hour out-of-service loss per spark plug case This case shows the need for improved cost figures before valid conclusions can be drawn.

In instances where one type of defect or irregularity is more serious than others, weighted defect values can be assigned to the individual irregularities represented in item 1. In this manner the curve will clearly and accurately portray the relationship between the hazardous condition and the time since overhaul.

During the past six months we have made considerable progress in the use of IBM equipment for compiling data and making calculations for drawing Overhaul Period Analysis Curves. Complete mechanization of this process will make it possible to do a hundred or more curves in the time now required to compile the data and make calculations for one or two curves.

Recently the U. S. Air Force has become very much interested in the potentialities of the Overhaul Period Analysis Curve as a management tool for improving the safety, reliability, and economy of operations. The AMC has been assigned the responsibility for studying and implementing the principles embodied therein.

DESIGN AND ANALYSIS OF TESTS

The use of statistical techniques in the design and analysis of tests has proven very valuable in designing, in analyzing, and in determining the significance of test results. Not only has this method reduced the number of test runs required and hence the cost, but also it has improved the reliability of the results. In addition, costly errors may be avoided. This technique has brought to light the fallacy of running a few tests and attempting to compare the results with mass data representing the present process. In addition, it has pointed out the value of running a test and a control in pairs where possible, subjecting both to the same environmental conditions. This method of designing tests has shown that in some cases the variables we were attempting to control, or to test, were not the only, or the most significant, variable in the process. In a properly designed test, two or more variables can be allowed to operate concurrently and the effect of each can be isolated and evaluated, and the interaction of the variables can be evaluated. We have come to the conclusion that every test, no matter how simple, should be viewed for proper test design and the results should be subjected to a statistical test of significance as the analysis of variance. This tool, unlike the analytical tools covered under the previous four (4) headings, is not a UAL development, but rather is covered in most text books on statistics. Some very excellent articles have appeared in recent issues of our "Industrial Quality Control" magazine. While our experience in the "Design and Analysis of Experiments" is still very limited, we feel that it is a powerful analytical tool that warrants wider use in the field of aircraft maintenance.

PERFORMANCE vs OVERHAUL TIME

A logical question to ask at this point is: "That's all fine in theory, but how has it worked out in actual practice?" Figure 17 shows the performance trend of our DC-6 aircraft for a seven year period, using the Irregular Removal Rate per Thousand Plane Hours Flown as the base for measuring performance. The first year of operation, 1947, has

DC-6 AIRCRAFT

PERFORMANCE vs. OVERHAUL TIMES

1947-1953

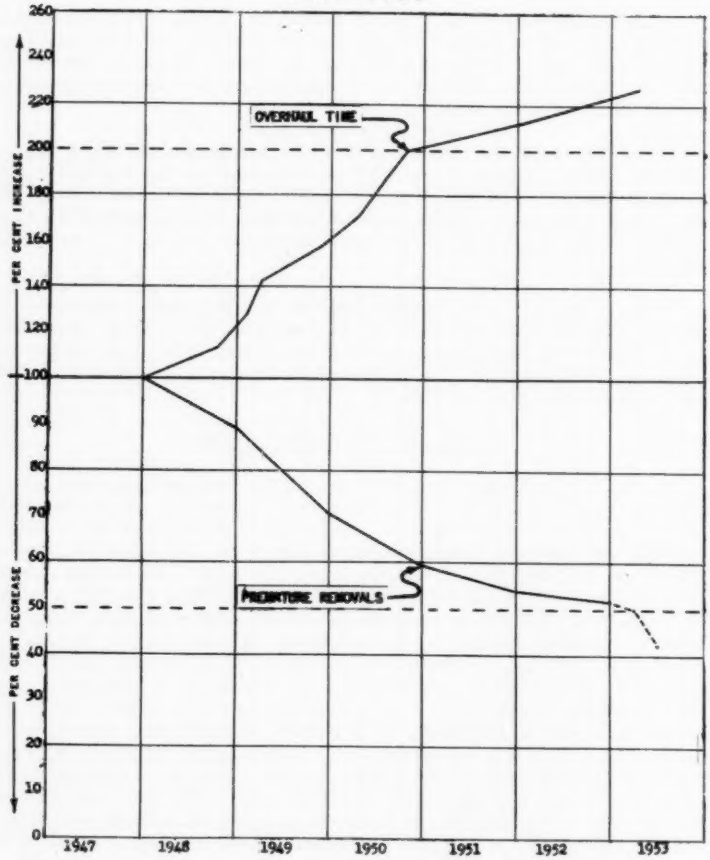


FIGURE 17

been represented as 100%, or par. Based on the 1947 Rate of Irregular Removal per Thousand Plane Hours Flown as 100%, the rate has been reduced to 50% - - the "failure" rate has been cut in half. This reduction in rate is the result of many factors, among which are training, experience, proficiency, improved techniques, methods, procedures, facilities, engineering modifications, maintenance analysis, operations research, and quality control. Furthermore, coupled with a 50% reduction in rate of irregularity is a 230% increase in overhaul time in DC-6 aircraft and engine components. Thus, thru the combined and cooperative effort of all activities, we have effected a four-fold gain. I believe the record speaks for itself, and we are only on the threshold of greater analytical techniques to come.

THE PAST, PRESENT, AND FUTURE OF QUALITY CONTROL IN PRECISION MANUFACTURING

Edward J. Schaller
Elgin National Watch Company, Lincoln Division

It is a privilege and a pleasure to be a guest speaker at our 8th Annual Statistical Quality Control Convention, in the beautiful city of St. Louis, the gateway to the South.

As some of you know, I had the opportunity of appearing at the 8th Midwest Conference at Davenport, Iowa last year. It is stimulating and encouraging to me to see the widespread interest and enthusiasm in quality control throughout the entire country.

The word statistics is like a red stop light to the average layman. Statistics have been used to confuse as well as to convince. The science of statistics is really no mystery.

Let us look at the past, the present, and the future in this field.

In a normal person's lifetime, there are on the average 32,648,000 precious minutes. We at Elgin are time merchants--the only way man has found to really measure his life's fulfillments.

Naturally, working with a product as important as this, quality is of primary importance to the product as well as to the consumer.

Return with me, if you will for a few moments, to the past, to the early 1930s, let us see the conditions that existed in American industry then.

Precision manufacturing was in operation during this period. On all critical close tolerance dimensions it was necessary to sort into "families" or "groups" parts of comparable known characteristics. After much time, they were assembled in a hit and miss method, and by process of elimination, we ended up with a satisfactory product. These were assembled by master craftsmen in the industry. Almost invariably these craftsmen were premium paid men that were actually sorting parts until one part was found that fit correctly.

But what transitions took place in the next years in industry! Wages, for one thing, increased 144% by 1950. This made it economically impossible to continue to produce goods at a price people could afford to pay. Working hours were shortened for the laborer, benefits increased, and the day of master craftsmen as production workers, was doomed.

Self-preservation of industry demanded new methods, new ideas, and a general decentralization of the skills of one to be broken down to semi-skills of many. Thus, the assembly line method of production was perfected.

This brings us to the present time. While the present assembly line production of precision parts solved temporarily the problem of self-preservation, it also created new problems. Foremost among these was the necessity for interchangeable parts. Unless this was perfected, the

effectiveness and economy of assembly line production would be minimized.

Quality control has played a major part in the accomplishment and improvement of this category. Now, more than ever before, we need to know what we can produce with the equipment we have, with the people we employ, and at what cost.

The Management of the Elgin National Watch Company has installed a very intensive program of Quality Control. Every employee at the Elgin Company is considered the most important person in the world. That fact, I believe, more than any other one thing, has been the reason for our outstanding results. After all, who can do more to make a better piece part than the worker himself?

Like most companies today, we have a union. Our experiences, with no exception, has been very gratifying in its acceptance of Quality Control. One important reason for this is the bond of confidence between the Union and ourself, in regard to Statistical Quality Control.

In no instance do we make a quality control installation without first discussing it with our supervision, our union steward in the department involved, and the employee or employees involved. This gives the individual an opportunity to discuss in private any and all objections to the proposed installation, as well as making him an active member in participating in the framework of the installation. His suggestions are most welcome. The results are a unified approach between the Company and the Union on a complex problem on which we both agree something must be done. Supervision carries the program from here. Each general foreman is provided with a notebook of current statistics on each operation in his department. (Figure 1).

The interest in the individual worker is carried still further.

Every chart, where possible, carries the employee's name in full and number. These charts are placed directly in front of the operators working area. In this way, he can compare the results of his efforts with his neighbor. The psychological effect is terrific.

Monthly conferences are held with the groups, with a representative of the various staff functions called in for counseling when necessary. Again, the union steward, as well as the foreman and general foreman attend. We compliment where possible, and offer constructive criticism and a basis for action when necessary.

It has been said by one of our operators that once she was indoctrinated with the principles of Quality Control she has never been quite the same again. This is true. Allow me to illustrate this experience on a most difficult job. This operation is known as the pivot polishing job on which the diameter of a pivot is held to plus or minus .0001 or about 15 times finer than a human hair. The job is performed by running a Ramet bar over the pivot for a mirror polished surface.

The diameter is determined by the amount of stock removed by the operator. In May of 1951, per cent defective control charts were started on this operation with an upper control limit of 10.2%. (Figure 2). By November, the upper control limit had dropped to 3.5% and in January of 1952, the upper control limit was 1.95%. This was the same operator on the same operation. Her individual work record was a stimulus to active

Figure 1.

ELGIN NATIONAL WATCH COMPANY - LINCOLN DIVISION

General Foreman's Quality Control Record

DEPARTMENT Train GROUP 4 PART NO. 489

OPERATOR & NO. --- SIZE & PART 19/0-1 Third Pinion OPERATION Final Inspection

TYPE OF CHART % Def. INSPECTION SPEC. Visual 100%

COMPUTATION	INSP. DATE		% INSPECTION				100% INSPECTION				TOTAL				CONTROL LIMITS	
TION DATE	From	To	No. lots	No. Pcs.	No. Rej.	% Rej.	No. Lots	No. Pieces	No. Rej.	% Rej.	Lots Mfg.	Pcs. Mfg.	Pcs. Rej.	% Rej.	Upper	Lower
4-23-52	12-17-51	4-21-52					184	240304	10494	4.4					6.0	
7-21-52	4-21-52	7-18-52					126	159831	7537	4.7					Fluctuation	
11-4-52	7-21-52	10-31-52					182	227043	8950	3.9					4.5	
1-27-53	11-3-52	1-23-53					76	92839	2542	2.7					4.1	
4-30-53	1-26-53	4-28-53					101	126988	3728	2.9					4.1	
8-7-53	4-29-53	7-31-53					94	121983	3389	2.8					4.1	
11-5-53	8-3-53	11-4-53					119	151818	4039	2.7					4.1	
2-10-54	11-5-53	2-8-54					116	133530	2538	1.9					3.0	

Figure 2

ELGIN NATIONAL WATCH COMPANY · LINCOLN DIV.

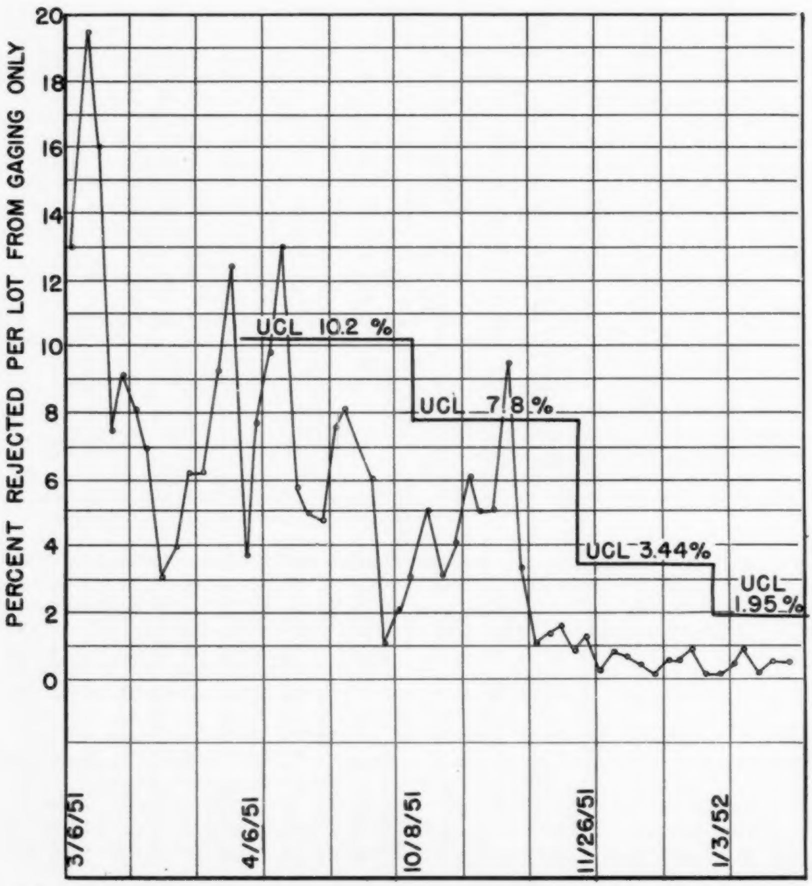
QUALITY CONTROL CHART

15/0-1F · PALLET ARBOR

JACOT FINISH STAFF

SALLY BALANCE 4069

TRAIN DEPT.



personal pride of workmanship. Competition by her neighbor had encouraged her to do better quality work and thereby increase her earnings. Certainly this operator will never be the same again.

Did you ever stop to think of why you work for the Company you do? Out of 10,000 people interviewed, by an independent research organization, the following reasons were given, listed in the sequence of importance to these individuals:

1. The reputation of the Company
2. The location of the Company
3. Security
4. Working Conditions
5. Pride of Workmanship
6. Money

We, in quality control, cannot do anything about the location of the Company. We can, however, by making the data available, enable the employee to do a better job, improve working conditions by method changes or equipment changes, and thereby decrease product cost which contributes to his security. We can, and do provide each employee with a picture record of his performance which enables him to constantly compare the results of his labors with his neighbors. Hence a pride of workmanship approach. Certainly we have, in many cases, provided the employee with additional earnings for less effort by these changes. All these things improve a Company's reputation in the community as well as among the employees.

Quality control is called on frequently by our Engineering Department to make a machine investigation to establish a basis for sound tolerances consistant with the economy of the operation. (Figure 3). The old method of setting a tolerance and hoping we could come close has become obsolete.

The Rate department calls on us frequently for data pertaining to a specific operation before establishing rates.

The Accounting department requests Statistical Quality Control data for cost purposes.

All departments use our services in some manner or other. We always attempt to fill any and all requests, for all we have to sell is service.

For those who are interested in establishing a Statistical Quality Control Department, Figure 4 shows our S.Q.C. organizational chart at our Lincoln Division.

I used to be of the opinion, at one time, that most workers were trying to see how much they could get from the Company and how little they could do to get it. Statistical Quality Control has convinced me that this condition does not exist. The average American worker is a conscientious and industrious individual with the same problems as you and I have, and the same pride of workmanship in his work as we have in ours.

For years we have labored to mechanize industry with more and better equipment, technical advances, and more economical methods. Billions of dollars have been spent for these things which have given the workers of the United States the highest standard of living in the world. Despite

Figure 3

ELGIN NATIONAL WATCH COMPANY LINCOLN DIVISION

MACHINE INVESTIGATION CHART

FLAT STEEL

GAGE ANGLE

15/0 COCK DOME

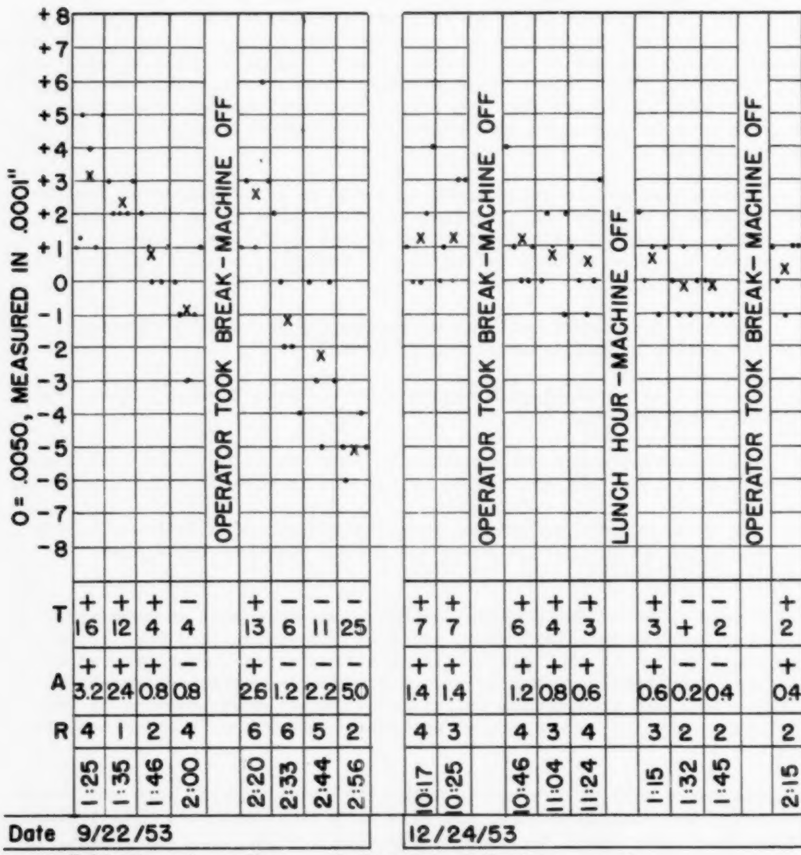
MACHINE 2

ORIGINAL INVESTIGATION

AIR EJECTION

FINAL INVESTIGATION

OIL FLOW ON CUTTER



• =Individual part readings X=Average of 5 piece sample

Figure 4

ORGANIZATIONAL CHART · LINCOLN DIVISION · QUALITY CONTROL

ELGIN NATIONAL WATCH COMPANY

predictions a decade ago, that machinery would cause widespread unemployment, the opposite has been true. More people are working in industry today than anytime in the history of our country.

But what about the people who run our machines? The American worker is striving for recognition by Management as never before -- and why not? Isn't the worker producing the part after all? Could you get along in your plant without them? Quality control helps give him the recognition he craves and deserves.

Occasionally you hear the old timer complaining to you that conditions were never like this in the "old days". That is true. Generally, however, the fellow who complains the loudest about the human relation problem today is usually the one who has helped create these problems in the past.

Let us use Quality Control techniques to explore this fertile area. I have heard many fantastic claims and figures expressed in dollars as to how much money was saved by using Quality Control.

How can you measure savings in Quality Control in dollars when the intangible value of how much you have influenced and affected the things that move an operator's mind and heart may have changed his work habits for the best interest of yourself and the Company for years to come? There is the real savings, in my opinion. Dollar savings are important. Psychological changes are also important.

Yet we, in Statistical Quality Control, have not scratched the surface. Space ships to the moon; atomic energy for heat, power and pleasure are all in this century. Let us hope and pray that all the new resources at our command will be used for the benefit rather than the destruction of mankind. We appreciate what the scientists, the statisticans and the universities have done for Quality Control advancement.

To the many missionaries who have given so freely and generously to the study, research and application of Quality Control, we say thank you. We can now photograph our results.

Quality control is the tool needed yesterday, available now for tomorrow and the problems ahead.

The wealth of America is in the quality of it's products.

INSTALLATION OF A QUALITY CONTROL SYSTEM

Warren E. Jones
Management Controls

Failure is a word that has seldom appeared in the Quality Control literature until recently and even then very little has been said or written about why modern Quality Control Systems fail. Success was the natural outlook for every program that was launched during the early years. No consideration was given to the possibilities of failure, of substantial cut-backs in QC personnel due to budgetary controls, and of management taking the gains from QC for granted. About nine years ago one of the leading Quality Control Directors who had already directed QC installations and systems for the previous ten years wisely said, "My most difficult problems today center around the justification of the QC budget. Most of our improvements in specification, scrap, inspection labor, and quality of product were accomplished more than three years ago; and, since these benefits are far removed from the current budget and today's pressing problem of reducing burden costs we are constantly under pressure to reduce our QC staff. Years ago I should have done more 'selling' of data and charts to continuously prove the value of our QC systems. My only practical recourse now is to request that our QC activity be entirely eliminated for a trial period in an entire department or plant to permit Top Managers to evaluate its loss." Reasons for QC failures should present valuable guide posts in making an Installation of a Quality Control System.

Top Managers who have discarded or curtailed QC all agree on one reason, "It's an Unprofitable system and it won't work in my plant!". Production Supervisors and Departmental Managers have mixed feelings. Most QC Engineers and Inspectors cannot understand why the "sharpest Management Tool in Half a Century" has been curtailed or discarded. They generally blame such action on lack of good judgment by the Top Managers or the Budget Committee; however, a cross-examination soon shows that the Top Managers were not given adequate training, reports and facts. Following are a few reasons why certain QC programs have failed:

1. Top Management backing has never been obtained. Most Top Managers who are expected to back a program for ultimate success do not know what Quality Control actually is; they have not been trained in the theory or the practice and are not in a position to evaluate its ultimate potential and to guide its expansion into maximum utility. QC has not been recognized and placed high enough in the organization to get corrective action results. This leads to high cost of QC in relation to the results accomplished.

2. Managers have been poorly trained, or have been severely criticized by the QC Program. They end up wanting QC eliminated for their own personal convenience or protection.

3. Quality Control Engineers who have made only limited application of the QC System on a few operations or to the receiving inspection will find their programs frequently curtailed. Potential application of QC in Specifications, Engineering Design, Methods, Standards, Sales, and Purchasing has seldom been tried. The Organizational set-up of most companies prevents the ordinary Quality Control Engineer from working

in or with these other Divisions. This gets back again to the basic problem of Top Management understanding, appreciation and support. The Quality Control Department will never find its proper prestige and position in most companies until selling and training are accomplished and only then will the full value of QC be obtained. Systems are being discarded because they were never "complete"; they may have done some good but they did not fully live up to the glowing expectations of Top Management.

4. Quality Control Directors have lacked the supervisory abilities to guide a good program through the stormy waters resulting in most plants when people are being evaluated and pressure is applied for corrective action. Practically every application of Quality Control passes judgment on human beings' ability to perform their jobs. Ordinarily Quality Controllers have always thought of evaluating the (a) machine, (b) method, (c) material, and (d) man. When the machine, method or material are causing trouble certain supervisory or staff employees are pointed out by the control charts for not having done their jobs correctly. The ideal Quality Control Director must really be a "super psychologist" to point out trouble, to evaluate people and to get something done about it without building up strong resentment to the QC system itself.

5. Quality Control Engineers have lacked "Ingenuity of Application". They have lacked the art of putting several sciences together to work profitably for their company. The sciences of numbers, mathematics, statistics, and probabilities are erroneously assumed to be sharp tools of management which will automatically work miracles when created. It is not realized that Quality Control Tools are as valueless as a sharp scalpel until it is in the hands of a "skilled" surgeon. Until QC Engineers become "skilled" surgeons using their QC tools profitably, they will never obtain the professional recognition they deserve and the support QC needs from Top Management.

6. Quality Control Engineers have not sold their Top Managers on the value of QC. Results must be converted into dollars saved wherever possible. These results should be confirmed by the general accounting records of the company and summary charts should be circulated monthly to constantly remind all concerned of the long range progress and value. QC Engineers have not sold shop supervisors on what QC will mean to them by aiding the company, they have kept results sheltered in their own offices or desks hoping some day that some one will ask for their records and do something about them. Good results have been obtained in spite of the negative personality of many QC Engineers. Every one must be sold on QC from Top Management on down to the Operators and all must get into the routine habit of thinking and acting on the basis of QC information and charts.

7. Quality Control Directors and Engineers have carelessly taken credit for improvements in a thoughtless manner that causes major upset within the organization. This upset occasionally

results in major set backs or elimination of QC itself. Many an excellent program has been curtailed immediately because it proved itself to be right when the assignable cause of trouble is the Top Manager who has jurisdiction over QC. Great care must be used when releasing embarrassing data. Occasionally it is good business for QC to lose a battle if that loss will strengthen QC's opportunity for winning the war. Many a good QC Engineer has been eliminated because he always proved that he was right and won every battle; he did not stay with the company long enough to win the war.

8. Quality Control Inspectors have been carelessly trained and sampling is not representative. Quality Control training methods are too complicated with considerable emphasis on the Mathematical & Statistical side of QC. Management believes QC is much too complicated for their plant.

9. Quality Control Engineers have poor statistical technique. Their paper work is much too complicated and the amount of sampling is not flexible. Initial Inspection cost to get processes under control can seldom be justified on a continuous basis to keep processes under control. There must be a planned program for simplification of paper work and reduction or increase of sampling.

10. Quality Control Inspectors and Engineers have assumed the Foreman or Design Engineer responsibilities resulting in positive and immediate elimination of QC when they make their first major mistake in these carelessly assumed capacities.

11. Quality Control Inspectors put on incentive pay arrangements concentrate only on getting complete coverage of every lot or process. This normal desire to earn more money usually makes the inspection of the sample less thorough and the control charts mean less than if fewer samplings were taken with absolutely thorough inspection. Incentive systems applied to Inspectors are undesirable and on occasion have caused QC Systems to be curtailed or discarded.

Other specific reasons for failure or set-back of QC Programs could be added to this list if my experiences were broader. It would be a pleasure to get letters from any Top Managers telling why they have curtailed QC or from any QC Engineer who has experienced curtailment, explaining why such action was taken in their particular case. Further information along this line could be published to guide future Installations of Quality Control Systems to insure complete application and lasting success in the eyes of Top Management.

Before starting the actual installation of a QC system, the desire for it must be in the minds of Top Management. Divisional Heads or Manufacturing Managers will have little trouble sewing these seeds of desire since their routine activities bring them into direct contact with Top Management. Staff supervisors such as Chief Inspectors, Design Supervisors, Methods or Standards Supervisors or Purchasing Agents will have to depend upon indirect contact with Top Management and work through their supervisors with literature, invitations to conferences or invitations to visit other plants. Many installations were started directly as a result of the enthusiastic suggestions of very young

members of the technical staff under this supervisory level. Lack of contact with Top Management is a major handicap resulting in lengthy delays before an all clear signal is given, a budget is provided and QC is formally recognized in the organizational structure.

A survey of the plant operations must be made before Top Management can be expected to pass judgment on its introduction. A positive request should be made for the time necessary to review the plant and to prepare sample control charts and QC studies on carefully selected operations to show how QC would work on a routine basis. Selection of these typical applications must be made in such a way that at least one or two will show some type of immediate corrective action during this short survey period. If any small amount of profitable corrective action is shown, Top Management will immediately see the potential and value of having this repeated many times daily on a routine basis. Emphasis must be placed on the fact that if many small improvements are made at the lowest levels in the plant, at the operator level or the set-up operator level, the composite effect will be substantial improvement in the entire plant. Middle or Top Levels of Management will in turn have fewer troubles and problems. To show how the Top Managers will get relief from major Quality problems will be sufficient to get their interest and their support for at least a trial budget and a chance to survey its value.

An opportunity to present QC to Top Management will eventually come. Be ready with your charts, facts, story of corrective action, plan for training, and plan for initial installation in the plant. It is generally wise to limit applications to one area in the plant; however, exceptions to this rule will be advisable if extremely profitable applications can be selected in each of several departments. Go slowly on the first actual applications, making certain that historical data is charted on each process before the control charts are put up and the routine control begins. Data before must generally be taken rapidly at a frequency that approaches continuous sampling otherwise corrective action might be made before enough data is obtained for comparison which will be used later to determine the benefits of QC.

Preliminary meetings are necessary to advise Supervisory groups of the general objects of the QC System. A preliminary meeting with the Union Committee is always advisable to gain their cooperation which is generally impossible if there is any mystery or uncertainty in their minds on any new program. When any worker on the job is clearly responsible for defective workmanship, a written warning should be given to the worker and penalties of some type inflicted after a fixed number of warnings are issued. Occasionally QC studies result in the workers request for a revision of the time standard on the job since "It can not be done that good without allowing more time on the operation!". Close cooperation is required during these time studies with a copy of the control chart on the job becoming a permanent part of the time study file. Good workmanship does not generally demand more time. For the few operations where more time is required the added cost will be more than offset by the increased productivity on other operations made easier and faster by more uniformity of quality on the previous operations. Before training is under way it is desirable to have one complete example from your own plant where the improvement of workmanship obviously helped productivity on a subsequent operation. This example must be hammered home to the Union Committee to show how total earnings will go up when QC is installed.

After these preliminary meetings, a positive program of training must be started. Material used during most of this training work should be taken from actual plant processes that tell a story of action with results. To get material for training a limited amount of regular control chart inspection must start immediately and be expanded as the training meetings continue. Training can not all be done within the plant to get the best results and it must not be concentrated within a short period of time. Specialized training must be planned for each of the following groups:

1. Top Management (President, VP's, Controller, Treas.)
2. Middle Management (Plant Manager, General Supt. etc.)
    a. Chief Engineer
    b. Methods & Standards Supt.
    c. Maintenance Supt.
    d. Purchasing Agents
    e. Accountants
    f. Personnel Director
    g. Sales Managers
    h. Chief Inspector
3. Lower Management (Department Supervisors)
4. Group Leaders and Set-Up Men
5. Machine Operators
6. QC Engineers
7. QC Inspectors and Bench Inspectors

The Top and Middle Management groups should be given preliminary intensive training at an executive day session. This first meeting must sum up what other companies are doing with QC, include instruction on the various types of charts to be used, demonstrate the basic type of sampling plans and the probabilities involved. Each executive must make at least one QC chart and understand the basic concepts of probabilities behind all QC records and analyses. Brief progress reports designed for review and further training are best included in the agenda of regularly scheduled meetings and should not be announced as training meetings. Each of the individuals in these groups should eventually attend one 8 to 10 day intensive Institute in Elementary Quality Control. It is essential that this training be removed from the plant operations and that it be scheduled over a long period of time; sending only one or two executives away each six months. All newly appointed executives should be scheduled to take one of these intensive courses which give uninterrupted training, enthusiasm and broader concepts of the ultimate potential of QC than can ever be obtained by in-plant classes. The wide variety of applications presented and association with executives from unrelated industries generally stimulates and suggests new and unusual applications.

The Lower Management Group should have five training meetings lasting two hours. This training will include practical visual demonstrations, preparation of the simplest type of control charts, and a practical demonstration of the risks and probabilities basic to QC. The first meeting should be a general summary of all the material to be covered. The next four meetings should be spaced from two to three weeks apart allowing time for experience to be gained in the shop, for the group to formulate questions and for them to encounter the operating problems arising with every new installation of a QC system. Encourage this group to participate in the refinement of the original procedures for installation and they will feel that they have had a part in its

creation.

Group Leaders and Set-Up Men should be trained in four or five hours, limiting their work to the visual demonstrations of sampling and the plotting of one average and range chart. All symbols and mathematics must be eliminated from these meetings. They should be limited to one hour each and deal only with problems in the departments and with how they are expected to make use of the system. They must be shown how QC will eventually make their jobs easier.

Machine Operators are seldom brought into formal training groups. The Lower Management Group is generally assigned the responsibility of explanation to the operators. Training in actual practice is generally done by memorandum ordering specific action to be taken by the operator when specific changes take place in the job. Occasionally operators must be shown the basic principles behind QC in order to overcome work habits that are not compatible to the success of a QC Installation.

QC Engineers must be carefully selected and then thoroughly trained in the QC technique. They must be selected wherever possible for their ability to get along with people. They should attend at least one elementary 8 or 10 day training course and eventually an advanced training course in QC. They should attend and participate in the activities of the American Society for Quality Control to keep up with new methods, forms etc. that are published or discussed.

QC Inspectors and Bench Inspectors should be carefully selected and then trained in the basic principles of sampling and shown how to make control charts. Training is generally limited to ten meetings of one hour each. Emphasis is placed on the need for representative sampling, accurate inspection of the sample, clear and precise recording and plotting of data, and firm but diplomatic enforcement of the QC procedures for corrective action when processes show lack of control. Visual aids are most effective with this group. Symbols and formulae of Statistics should be held to a minimum with emphasis on practical application within their own plant.

Rapid expansion is very desirable along with carefully written reports showing the "before" and "after" situation resulting from QC. A Quality Committee should be formed at the outset of the installation in order to spearhead the program and to insure its chance of survival through the initial test period and through its expansion into all operations or departments. Committee action can generally overcome difficult personalities which frequently are major stumbling blocks to a new QC system. This Quality Committee should be kept as small as possible but of necessity must Include Manufacturing, Methods, Maintenance, Purchasing, Design Engineering and Quality Control.

In conclusion, to make an Installation of a Quality Control System that will be "complete" and "stand the test of time":

1. Avoid every known pitfall resulting in failure,
2. Make a "survey" to plan the initial application,
3. Request that QC be recognized as a Department in the proper place in the organization so it will have a chance to prove its worth. It must be on an equal level with Manufacturing, Design, Sales & Purchasing.
4. Request adequate budget for manpower and expenses,

5. Plan preliminary meetings with great care.
6. Announce a specific and complete training program specifically designed for each different group that needs training.
7. Make your first routine applications with great caution, go slowly, insure success, be flexible and drop any initial study that does not show immediate promise of profitable results.
8. Expand your applications rapidly when the initial groundwork and applications have proven profitable.
9. Write a QC Procedure for cover each application and the duties of each member of the QC Department.
10. Streamline your paper work and reduce the amount and frequency of sampling under controlled conditions, making certain that these efficiencies are a matter of record.
11. Request the opportunity of concentrating QC analysis on every new machine, process or product that is brought into manufacturing.
12. Request preliminary or trial production runs before establishing critical design tolerances and the opportunity to recommend operating tolerances before the final design tolerances are established.
13. Make provision for follow-up and reporting of every Customer Complaint.

Sell QC with confidence and with the pride of a profession. Don't over-sell QC or promise results too fast. People always resist change and this resistance increases rapidly as the change is accelerated. It takes time to teach, to train and to get people in the habit of working the QC way. Make haste slowly, prove your results with caution, broaden your installation into all phases of your organization to make it really "complete" and your Installation of Quality Control will stand the "test of time".

References

This is a partial listing of articles or digests on Installation of Quality Control or the critically related subject of Quality Control Training

1. Halton, F. J., "How to Introduce Quality Control to Top Management", I Q C No. 6, Vol. II, 1946

2. Brumbaugh, Martin A., "Starting A Quality Control System" I Q C No. 6, Vol. II, 1946

3. Rahn, J. E., "Farmall Quality Control Training Program for Shop Supervision", I Q C No. 3, Vol. III, 1946

4. Thompson, E. D., "Some Methods for Selecting and Training Personnel in Statistical Quality Control", IQC No. 3, Vol. III, 1946

5. Wilks, S. S., "Research and Training in Industrial Statistics for the Future", I Q C No. 5, Vol. III, 1947

6. McHenry, W. K., "The Care and Feeding of a Quality Program" I Q C No. 1, Vol. V, 1948

7. Cohn, Alonzo C., "On a Quality Report to Management", I Q C No. 1, Vol. V, 1948

8. Weaver, Wade R., "The Foreman's View of Quality Control", I Q C No. 2, Vol. V, 1948; I Q C No. 3, Vol. V, 1948

9. Wiman, Charles D., "Quality Control Stimulates Thinking", I Q C No. 4, Vol. V, 1949

10. Robinson, E. H., "The Quality Control Department as a Training Ground for Future Supervisory Personnel", IQC No. 4, Vol. VI

11. Klock & Carter, "Quality Control in Six Minutes", I Q C No. 2, Vol. VII, 1950

12. Mundie, Paul J., "The Human Factor in Quality Control", I Q C No. 3, Vol. IX, 1952

13. Lifson, K. A., "Psychological Aspects of a Control Chart Installation", I Q C No. 5, Vol. IX, 1953

14. Juran, J. M., "Installing A Quality Control System", I Q C No. 7, Vol. IX, 1953

15. Clifford, Paul C., "Using Training Conferences as a Quality Control Catalyst", I Q C No. 7, Vol. IX, 1953

16. Wareham, R. E., "Methods of Training for Quality Control Engineers and Supervisors", p 331 Q C Conference Papers, 1953

FINDING AND MEASURING THE EFFECTS OF ASSIGNABLE CAUSES

Cuthbert Daniel
New York City

Most published quality-control stories have happy endings. The means-chart shows two successive points outside the three-sigma limits. "Investigation" reveals that the foreman in Bay IV was indisposed, or that the buffing-wheel on Machine 48 had some abrasive threads in it. Quite generally, one cause, or, occasionally, two concurrent causes are detected and the case is closed.

Unfortunately, the experience of this writer does not fit this pattern. Perhaps because of the enormous fees charged, or because such cases are by far the commonest, no cause can be found, even though out-of-control points continue to appear.

It is indeed natural to look for a single cause of an exceptional event, but it is also natural, when this search fails, to look for the multiple causes. Like other common-sense people, quality-control engineers make frequent use of Occam's Razor. The patent on this tool expired around A.D.1350, seventeen years after its successful introduction by William of Occam.

The razor is only a principle of applied logic. "Essentia non sunt multiplicanda praeter necessitatem" was the original statement. Roughly translated, William said - Don't increase the number of concepts you use, unnecessarily. Since the idea was invoked to cut down complexities, it was called Occam's Razor, or the Law of Parsimony. If one cause will do, then look no further. Parsimony, however, can be carried too far.

The causes of jumps out of control are either single or multiple. If single, they are either found or not found; if multiple, the effects of the several causes are either additive or non-additive. This paper will make no contribution to the search for singly-caused jumps. If they are found, good. If not, it may pay to look for multiple causes.

We will consider two main cases. In the first, each of several causes produces a small effect, not sufficient by itself to throw out of control the quality characteristic being measured. But the factors, or causes, are here assumed to act additively in the sense that if one factor moves the quality characteristic, say, 7 units, and the other moves it 5 units, then both acting together will move it 12 units.

In the second case the effects of factors are not additive. Thus two factors producing changes singly of 7 and 5 units, but operating non-additively, may produce a change of 18 units when they operate together. In statistical jargon, the factors are said to interact.

It will often be true that wider variation of the same factors will produce non-additive results even when the consequence of varying two or more factors a little is the sum of their individual effects. So do not try to simplify the picture by assuming that some factors always produce additive effects, or that some others invariably produce non-additive effects.

Since the factors we are looking for all have small effects, there is little chance of discovering them by inspection of single pieces of

past data. Nor is there much chance of finding them by deliberate variation of their levels in a test run or two, since the true effect of varying a single factor over a short range is generally smaller than the error standard deviation.

A more accurate means of analysing or collecting data must be found. Two such methods are available. They are called multiple regression (for the analysis of existing data), and fractional replication of factorial designs (for efficient collection of test data). These will be discussed, first for additive cases, and then for interacting cases.

We are often tempted to analyze the large amount of data that is produced in any plant using statistical quality control. It is natural to hope that voluminous data has a lot of information in it. This optimism is, in my experience, rarely justified. But what information is available may be useful in the design of plant tests. Since these experiments are usually expensive, all leads to be found in existing data should be exhausted.

The method of multiple regression has its own limitations, or assumptions.

a. The data must be a genuine sample of present conditions. Data taken when the process was different are useless, and it is not permissible to select favorable or unfavorable periods.

b. The error standard deviation of the effect or quality characteristic is the same for all combinations of levels of the factors (i.e. for all combinations of values of the independent variables).*

c. The levels of the factors are exactly, or nearly exactly known.**

d. The form of equation connecting the true values of the quality characteristic, Y', with the values of the k different factors, (X_1, X_2, - - - X_k), must represent the situation with a reasonable degree of accuracy. When the effects are small, the first equation to try is:

$$Y = b_o + b_1X_1 + b_2X_2 + - - - + b_kX_k \ . \qquad (1)$$

It is understood that the b_i and Y are only estimates of the unknown true values.

Suppose that we have from two to six factors (k is 2, 3, 4, 5, or 6), and further that we have about a hundred datum points. Each point

* If the standard deviation of the response, y, is a known function of the values of the factors, X_i, then a weighted multiple regression, al-somewhat more laborious, can still be carried out.

** If the values of the independent variables, X_i, are measured with error standard deviations σ_i , then the constants in equation (1) above will be measured with negative bias. It appears safe to use the spread of the observed X_i values, D_i^2, measured as a mean square deviation from the mean X_i value in the data as a yardstick. If σ_i^2/D_i^2 is less than 0.1, then it appears that the X_i can be taken as nearly exactly measured. See Hald (10) p. 616.

is an observed estimate of Y', called y, together with its corresponding k X_i-values. Suppose, finally, that assumptions a, b, and c, listed above are satisfied. Less than a week's work with a computing machine is then required to find whether or not the data contain any information on the linear, or first order, effects of the k factors on the true values Y'. Notice that I did not say that it is possible to find the k slopes, or effects, b_i. The data may not have this information in them in separable form.

Three reasons seem to account for most of the failures of such analyses to produce significant b_i. First, the independent variables, X_i, may themselves be closely correlated in the plant's operation. The data will generally show if this is the case. Second, the wrong factors may have been chosen, the process being tolerant of variation in those actually used. Third, a linear equation may not be adequate to describe the true situation.

Up to this point I have not discussed the key element in the whole procedure. Its omission is by far the commonest reason for the failure of this type of curve-fitting. This requisite is, bluntly stated: the analyst must understand multiple regression. Mere ability to take punishment is not enough. In almost every company someone has worked out his own "calculation sheet" for multiple regression or multiple correlation, but if my own experience is typical, half of these sheets are worthless. Or worse.

Mr. Gauss and Mr. Doolittle are often given credit for devices they never heard of, and for errors they never made. Most common of all, one of Gauss' major contributions is not mentioned. A regression analysis is usually of little value if it does not include estimates of the standard errors of the b_i. Gauss showed how these can be found, and all good statistical texts give his method. As examples, I would cite Hald (10), and Snedecor (13). Engineers will, I believe, find Hald easier to read than Snedecor.

When a regression analysis is successful, several of the b_i are measured with acceptable precision. These "regression coefficients" or slopes, are found to predict effects, - individually small for the usual range of their corresponding X_i - which will occasionally suffice to pile up a Y value which will lie outside the control limits.

Some quality control engineers may respond to such a finding by deciding to put control limits on the influential X_i. They will, of course, study the pattern of correlations between the X_i to avoid using too many new control charts.

But it may be impractical to get some X_i into a state of statistical control, and sometimes it is not necessary. We can use knowledge gained from the successful fit of equation (1). One important factor can be made to buck another. Thus, when the temperature increases by 2 degrees, it may be practical to reduce the pressure by 0.72 p.s.i., these numbers being in the inverse ratio of the b_i's found for these two factors. In this way we can sometimes enforce correlations on some of the disturbing X_i so that their combined effect is small.

After a successful regression analysis it will usually be necessary to prove that the factors found continue to have the effects predicted. This can be done only by deliberately varying the factors to

see if they produce the predicted effects. In many cases, this can be done directly in the plant, without producing any defective material. The most efficient way now known is by a series of runs that are balanced like a "multiple-factor orthogonal design of first order."

It should be added that a regression analysis is sometimes successful even though it turns up no significant b_i. This may mean that several factors may be eliminated as effective elements. When this is the case, then other factors must be varied in the plant tests.

In summary, the principal use of the analysis of previously collected data is to aid in selecting the factors and levels to be studied in a planned sequence of controlled runs.

Many of you are familiar with the three-factor design of first order. Calling the three factors A, B, and C, and indicating the two levels of each factor by the symbols 0 and 1, the four runs required are shown in Table 1.

Table 1

Three factors, each at two levels

Run No.	A	B	C
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	0

When the effects of the three factors are additive over the range of variation used, then the results of the four runs can be assembled in three ways, so as to measure the effect of changing each factor as precisely as if the whole set of four had been devoted to the study of one factor.

But it is with the extensions of this design that we are concerned here. For example, if as many as seven factors are suspected, and if they can be independently varied over ranges like those that occur in the plant, then Table 2 gives a sequence of eight runs that offer considerable advantages.

Table 2

Seven factors, each at two levels

Run No.	A	B	C	D	E	F	G	Run No.	A	B	C	D	E	F	G
1	0	0	0	1	1	1	0	5	1	0	0	0	0	1	1
2	0	0	1	1	0	0	1	6	1	0	1	0	1	0	0
3	0	1	0	0	1	0	1	7	1	1	0	1	0	0	0
4	0	1	1	0	0	1	0	8	1	1	1	1	1	1	1

Seven conclusions can now be drawn, each as precise as if it alone were under study. Since the error standard deviation, σ_o , is usually known in moderately well-controlled processes, we can calculate the 95% confidence range on each effect, before we make the runs. The range will be $\pm 1.96\ (\frac{1}{4} + \frac{1}{4})^{\frac{1}{2}}\ \sigma_o$, or $\pm 1.39\ \sigma_o$ for this experiment.

Another advantage of this type of plant test is that the variations in quality of product introduced into single runs are generally so small that defective material will not be produced. This aspect of factorial designs (and of their fractional replications) has not been sufficiently emphasized. It gives a new meaning to the term Factorial Designs. They are designs that can be carried out in a factory.

Table 3 gives the basic design for studying as many as 15 factors, each at two levels, provided only that their effects are additive. The 95% confidence range on each effect is $\pm 0.98\ \sigma_o$. This design is, then, the smallest one that can be relied upon to detect effects of the order of one standard deviation, and even this one will miss half those effects of magnitude σ_o . Plackett and Burman (12) give designs for as many as 99 factors.

Table 3

Fifteen factors, each at two levels

Run No.	A	B	C	D	E	F	G	H	J	K	L	M	N	O	P
1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0
2	1	1	1	0	0	0	1	0	1	1	1	0	0	0	1
3	1	1	0	1	0	1	0	1	0	1	0	1	0	0	1
4	1	1	0	0	0	1	1	1	1	0	0	0	1	1	0
5	1	0	1	1	1	0	0	1	1	0	0	0	1	0	1
6	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0
7	1	0	0	1	1	1	0	0	1	1	1	0	0	1	0
8	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
9	0	1	1	1	1	1	1	0	0	0	0	0	0	1	1
10	0	1	1	0	1	1	0	0	1	1	0	1	1	0	0
11	0	1	0	1	1	0	1	1	0	1	1	0	1	0	0
12	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
13	0	0	1	1	0	1	1	1	1	0	1	1	0	0	0
14	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
15	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Designs for factors at four levels are also available and all the intermediate cases are easy to construct - with any number of factors from five to none at four levels, the remaining 15 degrees of freedom being used for two-level factors. Similar designs, called orthogonal squares, are given in Fisher and Yates' Tables (9), and in Cochran and Cox (3), for studying 4, 5, 6, 8, 9, and 10 factors, each at 3, 4, 5, 7, 8, and 9 levels in the respective designs.

When such a plan is carried out, with the usual precautions as to randomization, the results can be depended on with a surety not possible

with the results of a regression analysis.

We proceed now to the second broad class of situations in which a single assignable cause can not be found. These are the non-additive situations: the effect of A and of B operating together is greater than (or less than) the sum of their effects operating singly.

Probably the simplest way to mine existing data for information on the non-additive effects of, say, k factors, is to try fitting an equation of the form of (2), to about 100 points that satisfy all three assumptions discussed in the section on linear multiple regression.

$$Y = b_0 + b_1X_1 + b_2X_2 + \quad --+ b_kX_k + b_{12}X_1X_2$$

$$+ b_{13}X_1X_3 + ---- + b_{k-1,k}X_{k-1}X_k \qquad (2)$$

For 2, 3, 4, and 5 factors, there will be 4, 7, 11, and 16 constants to be evaluated on the right side of equation (2). The first two of these cases are quite manageable with a desk calculator. The latter two are not at all difficult if IBM 602 or equivalent equipment is available.

If such an equation does not give a satisfactory fit, the next attempt would probably be to include the k "square terms",

$$X_1^2, X_2^2 \quad --- X_k^2 .$$

The numbers of constants to be estimated are 6, 10, 15, and 21, for the 2, 3, 4, and 5 factor cases respectively, and of these, only the first is practical without automatic machinery.

The recommendation that you understand the multiple regression theory before you apply it, is still in order.

In favorable cases, a regression will suffice to rule out some factors, and to point more or less clearly to the importance of others. Again, the next step will usually be to conduct a series of plant or pilot-scale test runs to verify the findings of the analysis. It will always be necessary to plan the sequence so that all interactions revealed or suggested by the regression analysis can be measured separately.

The appropriate experimental plans are called Fractional Replicates of Factorial Designs. These plans were first discussed by Finney (7, 8), and by Kempthorne (11). There is an explanation in Cochran and Cox (3), pp. 189-194, and an excellent introductory article by Davies and Hay (5). The writer has found the paper by Brownlee, Kelly, and Loraine (1) most useful for selection of an appropriate plan. Examples of the actual use of such designs have been given by Del Priore (6), by Brownlee (2), and by Daniel and Riblett (4).

Since the objective is to draw conclusions about the effects of one particular factor <u>at each level</u> of another factor, a larger number of runs is generally required. Roughly speaking, if N runs are made, instead of N-1 conclusions about main effects, we can now count on from N/3 to N/2 conclusions about main effects, <u>and</u> about the various possible kinds of non-additivity. Here, too, each conclusion is of maximum

precision, using all N pieces of data, and so, at least for two-level factors, further improvement does not seem imminent.

Examples of these designs will not be given here. A review of designs found useful, and of simplified methods for their construction, was given at the Ithaca meeting of the Institute of Mathematical Statistics, on March 20, 1954.

In summary, two common reasons for the failure to find assignable causes are given: either several factors operating additively are producing lack of control, or several factors operating non-additively may be blamed. An outline of how to analyze past data is offered and some suggestions are made for the efficient collection of test data. The two statistical tools required are multiple regression and fractional replication of factorial designs.

I wish to acknowledge the editorial assistance of J. A. Maxwell in preparing this paper.

REFERENCES

1. Brownlee, K. A., Kelly, B. K., and Loraine, P.K., Biometrika, 35, 268-276, 1948.

2. Brownlee, K. A., Annals of N.Y. Acad. of Sciences, 52, 820-826, 1950.

3. Cochran, W. G., and Cox, G. M., Experimental Designs, Wiley, 1950.

4. Daniel, C., and Riblett, E. W., A Multifactor Experiment, accepted for publication, Ind. and Eng. Chemistry, 1954.

5. Davies, O.L., and Hay, W. A., Biometrics, 6, 233-249, 1950.

6. Del Priore, F. R., and Kommers, W. J., An Example of a Fractional Replication in a Bearing Abrasive Wear Test, Paper read at annual meeting of Amer. Statist. Assoc., Chicago, December 1953.

7. Finney, D. J., Annals of Eugenics, 12, 291-301, London, 1945.

8. Finney, D. J., Jl. Agricultural Science, 36, 184-191, 1946.

9. Fisher, R.A., and Yates, F., Statistical Tables, Ed. III, Hafner, N. Y. 1949.

10. Hald, A., Statistical Theory with Engineering Applications, Wiley, N. Y. 1952.

11. Kempthorne, O., Design and Analysis of Experiments, Wiley 1950.

12. Plackett, R.L., and Burman, J.P., Biometrika, 33, 305-325, 1946.

13. Snedecor, G.W., Statistical Methods, Ed.IV, Collegiate Press, Ames, Iowa 1946.

STATISTICAL CONTROL OF COMPLEX PROCESSES

James V. Strela
Thompson Products, Inc., Cleveland, Ohio

In this paper, consideration is given to practical statistical techniques applicable to the control of measurable characteristics resulting from repetitive processes of varying complexity. By statistical control we mean the variability pattern of measurable characteristics resulting from the operation of the unavoidable chance causes of variation as contrasted with the additional variability resulting from the operation of the detectable assignable causes which are subject to economical detection and elimination.

While the concept of statistical control is generally well understood and may be defined, the definition of a complex process in apposition to a simple process defies precision of definition. For simplicity and complexity are relative terms and must always be thought of in terms of the conditions associated with the process. To use an analogy, to Albert Einstein the theory or relativity is clear and perhaps basically simple while to most people it is complex beyond comprehension. Similarly, producing quality characteristics conforming to engineering specifications may be easily attainable on one set of simple machines but very difficult on the same set of machines after they are worn out or when operated by unskilled or careless operators. But even a skilled operator running the same machine and using basically uniform material may produce various strata of quality from day to day if he has no technique to guide him in his effort to meet the desired standard of quality.

It should be obvious that if time alone may be a factor contributing to the varying quality strata produced by simple machining or other operations, the complexity of the machine, such as the multiple spindle screw machine, multiple drill press, multiple grinder, etc., is likely to present additional problems in producing values conforming to quality specifications. While the fact stated is generally known, because industry is constantly coping with it, the statistical techniques applied to such processes are not always most appropriate for the control purposes.

Some methods, such as the analysis of variance, difference between two or several means, test of significance, correlation analysis, etc., maybe recognized by the statistically minded Quality Control Engineer as ideally suitable from the statistical viewpoint for the statistical analysis of such processes. However, they may be poorly suitable from the administrative standpoint so that a practical approach, even if less precise, is usually sought. As this paper is written for such practical purposes, only the basic and easily administered techniques are here considered. Furthermore, since the fundamental basis of those techniques is pretty generally known to the older members of the ASQC, this paper will emphasize the logical and statistical assumptions on which the techniques are based rather than case histories which if presented without the underlying theory and philosophy might prove unconvincing. (However, if time permits, case histories will be presented on slides in support of the theory developed and demonstrated by reference to an artificial process with various manipulations to simulate either simple or compound processes in which we are interested).

Let us now procede with the development of the subject by first considering the determination of a statistical standard of what may be called a simple controlled process represented by a large distribution of N values from the Thompson sampling machine (Figure 5). However, from this point on, we forget the machine and think of the values produced by it. The values are presented, analyzed, and interpreted in relation to an assumed quality standard like values produced by innumerable industrial simple or compound processes simulated by the operation of this teaching device.

Fig. 1. Relation of Sample Size to Reliability of Standard Determination

In this figure, an attempt is made to demonstrate that a controlled process, in the long run, distributes normally, and that the sample size bears a relation to the reliable estimate of that pattern. But first of all, the figure is to illustrate two ways of determining a practical standard; one we may call the undisputed evidence presented by the face value of the sample; the other, the more economical one, we may call the statistical method.

The first method simply relies on the evidence of the sample values, without any statistical analysis. The four samples, charted as the four modified bar diagrams, represent 5, 50, 500, and 5000 values from the sampling machine when operated on a horizontal plane, thus producing a single quality stratum, represented by its pattern of variability and its position on the scale of values.

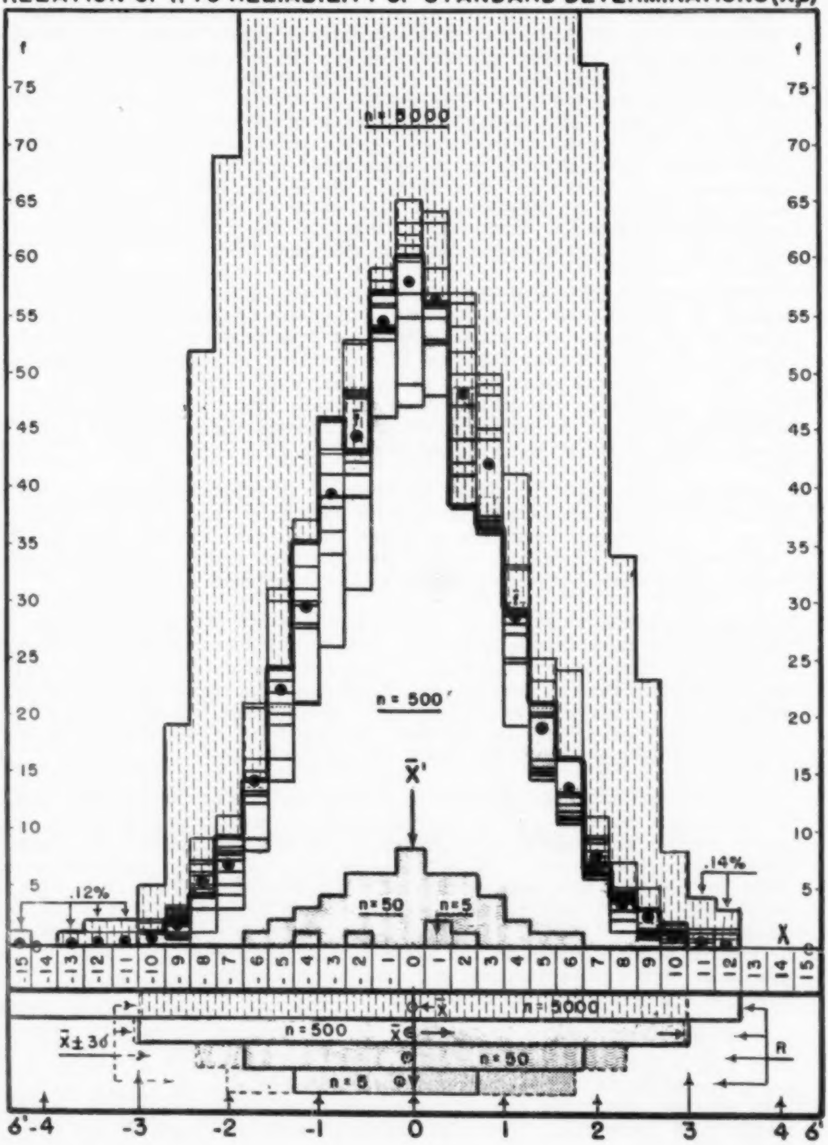
Obviously, if the four diagrams are representative of a sampling process, we can infer that the more we sample the more the values spread and that even 5000 values may not present the full picture of the distribution pattern of the whole process. However, if the sample of 5000 values revealed only 6 and 7 values beyond the range of the sample of only 500 values, we may conclude that, for practical purposes, if we know within what spread and pattern 99.73% of the values produced are distributed, we know practically all there is to be known about that process; consequently, we may consider that pattern and its location on the scale of values a realistic enough standard. By standard, we mean a representative mean ($\bar{X}'$) and standard deviation (σ') of the theoretically infinite number of values produced by the process. As the figure shows, the sample of five, or even of fifty values, is hardly a sufficient basis for the determination of that standard, unless something is added to the bare skeleton of such samples.

This brings us to the second method of the standard determination, which is statistical. In the figure, it is represented by the horizontal bars in the bottom section of the figure. The bars bordered by the solid vertical lines represent the ranges (R) of the four samples; the hollow circles with a dot inside them represent the four sample means ($\bar{X}$); the dotted vertical lines represent the mean plus minus three standard deviations ($\bar{X} \pm 3\,\sigma$) of the four samples as the four estimates of the probable process spread within which 99.73% of its values are located.

Note that in the two smallest samples the dispersion of the sample ranges is much smaller than the estimated $\bar{X} \pm 3\,\sigma$ spread of the two samples. Let this example be taken as an illustration of the additional information provided by the statistical analysis. Note also that the $\bar{X} \pm 3\,\sigma$ of the sample of 500 and of the 5000 values are practically identical. This fact may be taken as evidence that sampling beyond the

Fig. I.

RELATION OF n TO RELIABILITY OF STANDARD DETERMINATIONS ($\bar{X}'\sigma'$)



500 values from this process is practically a waste of time.

The figure emphasizes the relationship between the reliability of the standard determination and the sample size. That reliability increases as the sample size increases, but beyond a certain point that additional reliability is of little practical significance. If, therefore we assume a quality specification for the process represented to be 0 ± 10.5, we may say the process is capable of meeting that specification except for the .26% of values beyond its limits which, however, may be considered an acceptable percentage of substandard quality. In short, the $\bar{X} \pm 3\sigma$ of this process, within which there are 99.74% of the 5000 values just analyzed, is now considered the true process standard described by $\bar{X} = \bar{X}' = 0$; $\sigma = \sigma' = 3.5$. Naturally, in this standard, the mean value $\bar{X}' = 0$ represents a definite value X, such as the desired specification mean $\bar{X}'_{sp} = 2.000''$, 350 psi, 10 lbs., 60 cc, etc., depending on the type of the characteristic measured.

The solid dots in the figure represent the cell frequencies in the sample of 5000 values but drawn to the scale ten times smaller than that used in the figure. The shaded figure is a perfectly normal distribution based on the established standard $\bar{X}' = 0$, $\sigma' = 3.5$ based on the sample of 5000 values as a proof that, in the long run, this process distributes normally. The distribution of these 5000 values practically coincides with the normal distribution pictured; there are only three class interval deviations ranging from 10 to 30 units.

We are very much interested in proving the existence of this normal distribution pattern because we wish to make it a basis for the technique of the control chart for averages (or totals) and ranges (or standard deviations) of various sample sizes to test the conformance of the values subsequently produced to a previous standard.

The statistical properties of the normal distribution may be described by the heavy concentration of values about its central or average value ($\bar{X}'$). The rate of decrease of the frequencies from this central value is irregular but continuous. This gives the distribution its bell shape. As a result, we may infer the following laws concerning the distribution pattern of the averages and ranges of n values such as are used in the control chart technique subsequently used:

1) Since the normal distribution is symmetrical, the random sample n values will also be more or less symmetrical about the $\bar{X}'$. While small samples must deviate considerably from the full pattern of their parent population, their combined average will be a very good representation of the true but unknown mean of the individual values of the whole population $\bar{X}'$ and, even in shape, the combined large sample will conform closely to the pattern of its population. This inference is confirmed by the close conformance of the distribution pattern of the large sample of 5000 values pictured and the shaded normal distribution based on $N = 5000$, $\bar{X}' = 0$, $\sigma' = 3.5$.

2) Since the normal distribution is symmetrical and bell shaped, most sample random values will be from the central region of their parent population; as a result, the averages ($\bar{X}$) of such samples will also cluster about the mean of the parent population, especially as the sample size increases. The relationship between the sigma of the individual values (σ_X) and the sigma of averages of the n values ($\sigma_{\bar{X}}$) is this: The latter is the former divided by the square root of the sample

size ($\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$).

3) Since the normal distribution is bell-shaped, the larger the sample the greater the probability for the rarer values to appear in the sample. The larger the sample, the larger its range (or standard deviation) as is well illustrated by an increasing range of the four samples in the figure. The sigma of such ranges (σ_R) also increases with the sample size. This relationship may be simply stated by saying that the sigma of the ranges of n values is k times the sigma of the individual values of their universe ($\sigma_R = k\sigma'$). The size of the factor k depends on the sample size while the sigma of the population, even if unknown, is stable as long as the population itself is stable.

4) In the long run, the average range ($\bar{R}$) of a large series of m samples of n values from a stable process becomes a stable measure of variability of such n values. Now, since the sigma of the population (σ') is another constant, the ratio between the representative average range ($\bar{R} = \bar{R}'$), as a constant, and the sigma of the universe, as another constant, can be only another constant, the magnitude of which depends on the sample size. This constant is described by the symbol d_2 ($\bar{R} / \sigma' = d_2$).

5) Since the ratio between the average range and the sigma of the population is the value of the factor d_2 just mentioned, if only the average range of the m samples of n values and the corresponding value of the factor d_2 are known, the unknown sigma of the processes or population may be estimated as the ratio between the average range and the factor d_2 ($\sigma'_r = \bar{R}/d_2$).

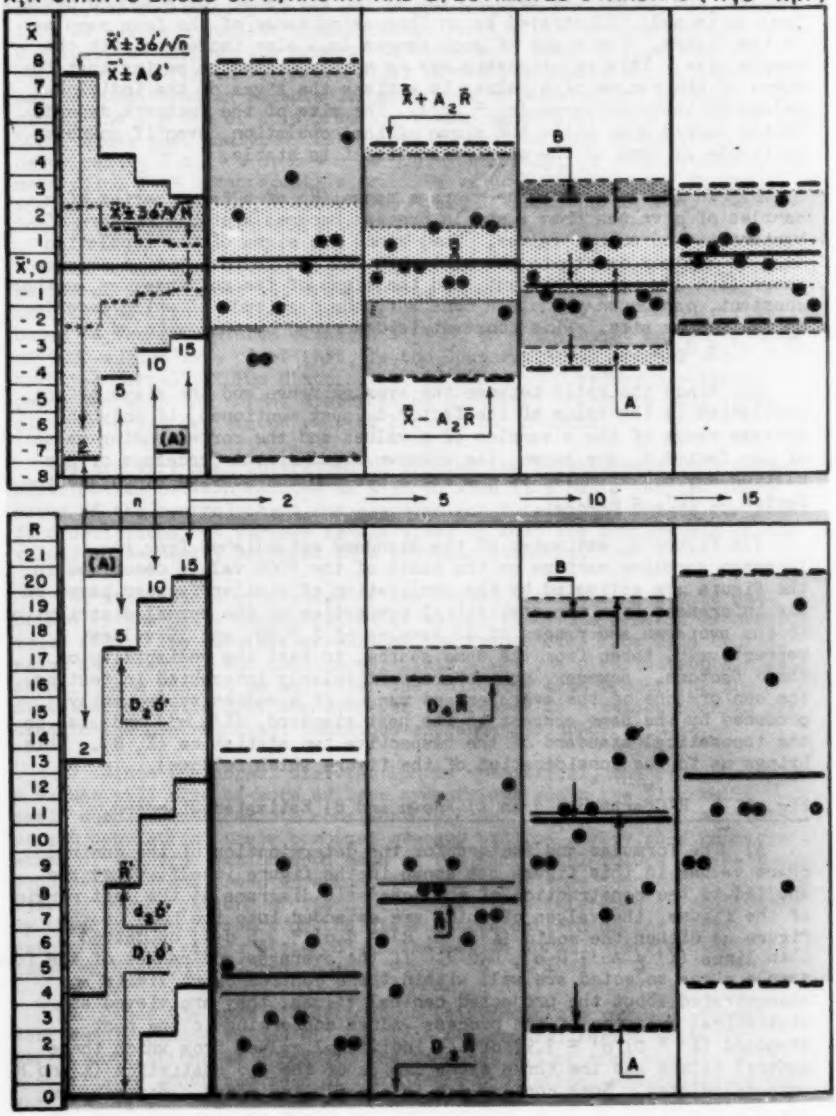
(In Figure 4, estimates of the standard established from the Thompson sampling machine on the basis of the 5000 values described in the figure are estimated by the application of similar factors based on the inferences from the statistical properties of the normal distribution to the averages and ranges of 10 samples of 2,5,10, and 15 values, respectively, taken from the same source, to test the reliability of those factors. However, now we are particularly interested in testing the conformance of the averages and ranges of n values subsequently produced by the same process to its past standard, ($\bar{X}'$, σ') and also to the theoretical standard of the respective two statistics ($\bar{X}$, R). This brings us to the consideration of the figure which follows).

Fig. 2. $\bar{X}$, R Charts Based on A) Known and B) Estimated Standard.

A) The formulas and factors for the determination of the control chart values in this figure are shown in the figure itself. They are applied to the construction of the schematic diagrams at the left margin of the figure, the values of which are extended into the body of the figure as either the solid ($\bar{X}' = 0$; $\bar{R}' = d_2\sigma' = 3.5\, d_2$), or dotted thin lines ($\bar{X}' \pm A\sigma'$; $D_1\sigma'$, $D_2\sigma'$). If the averages and ranges of the four sample sizes selected are well within these control chart limits and concentrated about the projected central values, they are viewed as statistical evidence of the process values conforming to the past standard ($\bar{X}' = 0$; $\sigma' = 3.5$) of the individual values from which these central values and the three sigma limits of the two statistics ($\bar{X}$ and R) were calculated. Such conformance is clearly indicated. However, we are not surprised at this conformance; for these samples are from the same source as the 5000 values previously analyzed and from which the original standard ($\bar{X}' \pm 3\,\sigma' = 0 \pm 10.5$) was established. Weare chiefly

Fig. 2.

$\bar{X}$, R CHARTS BASED ON A) KNOWN AND B) ESTIMATED STANDARD ($\bar{X}'$, σ' - $\bar{\bar{X}}$, $\bar{R}$)



interested in demonstrating the reliability of the method applied to this experiment.

Now in shop practice we do exactly the same as we did in this case; for once we determine a standard, we wish to maintain or improve it. When the standard is impaired, we wish to detect such deterioration as soon as possible. The method illustrated serves that purpose.

B) Often, however, we do not have sufficient time to determine a standard from a huge sample. We must then estimate it from whatever we have in hand. From such evidence, we must also determine the state of control. The average range ($\bar{R}$) and the grand average ($\bar{\bar{X}}$) of the m samples such as are shown in the figure as the solid dots are then the basis for such determinations. The two central values thus established are then viewed as the best available estimates of the true parameters of the two statistics ($\bar{X}'$, $\bar{R}'$). In the figure, the heavily drawn solid lines and dotted lines about them represent such estimates, based on very limited evidence as compared with the statistics based on 5000 values (thin lines). The state of control is again indicated; for all 40 averages and ranges plotted are well within their respective control chart limits, actually clustering about them.

Fig. 3. Comparison of True Standard ($\bar{X}' \pm 3\,\sigma'$) and Its Available Estimates ($\bar{X} \pm 3\,\sigma$)

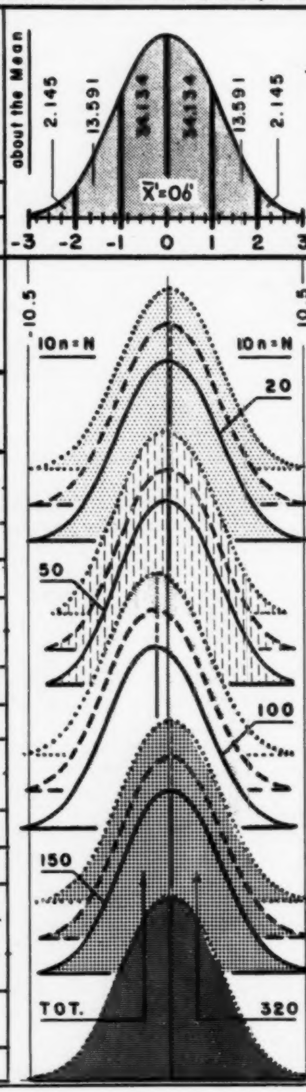
When we know the process is in control, from the grand average ($\bar{\bar{X}}$) and the average range ($\bar{R}$) of the m samples of n values such as illustrated in the previous figure, we may estimate within a reasonably small error the unknown distribution pattern of the whole process. This is done in Figure 3. The true standard established from the 5000 values is recapitulated on the top while the twelve estimates of the same "supposedly" unknown standard, based on the $\bar{X}$ and $\bar{R}$ or the four sets of 10 samples of the four sizes selected and charted in the previous figure, are drawn to scale. The schematic curve on the bottom is the thirteenth estimate based on the combined 320 values.

The table in the figure presents the formulas used for the three methods of estimation (σ, $\sigma_r' = \bar{R}\, d_2$, $\sigma_s' = \bar{\sigma}/c_2$); the necessary factor values are given in most quality control texts. At this point, it is to be stressed that the percentages of the values within the assumed quality specification, 0 ± 10.5, which is made to correspond to the $\bar{X}' \pm 3\,\sigma'$ or the natural limits of variability of this process, agree within a few tenths of one percent. This much, then, for the proof of the reliability of the technique presented in determining the unknown standard from very limited evidence obtained from a controlled process.

The analysis charted in Figure 3 shows that sampling may be economical as well as reliable in describing the unknown stable process. We shall now consider a process which, however, is not so simple or stable as the process just analyzed; for that reason, we shall endeavor to infer a sampling method best suitable for the detection of the undesirable elements. The theoretical background for such inferences is presented in the next figure.

Fig. 3.

COMPARISON OF TRUE STANDARD ($\bar{X}' \pm 3\sigma'$) AND ITS AVAILABLE ESTIMATES ($\bar{X} \pm 3\sigma$)



% Areas of the Normal Curve within Zones of 3 sigmas	TRUE STANDARD (BASED ON MANY THOUSAND VALUES FROM THE THOMPSON SAMPLING MACHINE WHEN OPERATED ON THE HORIZONTAL PLANE)			% Within the $\bar{X}' \pm 3\sigma'$ of a Controlled Process = TRUE STANDARD =	= 100% or BASE for the purpose of calculating the % of Standard
$\bar{X}' - 3\sigma'$	$\bar{X}' + 3\sigma'$		$1\sigma' = \sqrt{\Sigma(X-\bar{X})^2/N}$		
-10.50	10.50		3.50	99.73	100.00

12 ESTIMATES OF $\bar{X}' \pm 3\sigma'$ BASED ON 4 SETS OF 10 SAMPLES OF 2, 5, 10, & 15 VALUES X FROM T. P. QUINCUNX

	Estimate of $\bar{X} - 3\sigma$	(1)(2)(3) σ $\bar{X}+3\sigma$	↓	σ $\sqrt{\bar{\sigma}^2}$	σ'_r $\bar{R}/d_2$	σ'_s $\bar{\sigma}/c_2$	% WITHIN $\bar{X}' \pm 3\sigma'$ (0±10.5)	% OF STANDARD (99.73)
	-10.55	11.05	(1)	3.60			99.64	99.91
1.	-10.64	11.14	(2)		3.63		99.61	99.88
	-10.64	11.14	(3)			3.63	99.61	99.88
				(1)	(2)	(3)		
	- 8.84	9.28	(1)	3.02			99.95	100.27
2.	- 9.59	10.03	(2)		3.27		99.87	100.14
	- 9.53	9.97	(3)			3.25	99.88	100.15
	-11.12	9.76	(1)	3.48			99.69	99.96
3.	-10.91	9.55	(2)		3.41	3.41	99.76	100.03
	-11.39	11.15	(3)			3.57	99.62	99.89
	-10.39	11.15	(1)	3.59			99.23	99.50
4.	-10.15	10.91	(2)		3.51		99.23	99.50
	-10.24	11.00	(3)			3.54	99.18	99.45
	-10.50	10.50	(1)	3.50			99.73	100.00

Fig. 4. Adverse Effect of the Continuous Shift of the Process Components on an Assumed Quality Specification.

In this figure, consideration is given to the problem of quality stratification resulting from such factors as tool wear, tool slip, tool chip, changing job set-ups, or simply from the fact that whenever a quality characteristic is produced by several operators, on the same or different machines, or at different times, the problem of stratification enters in. In the previous examples analyzed, that problem was ignored because the characteristic studied represented a single stratum of quality ranging within the $\bar{X}' \pm 3\sigma' = 0 \pm 10.5$ resulting from the stability of our sampling machine. However, industrial operations are not so stable as that machine when it is operated on the horizontal plane. Now consideration to a changing process must be given.

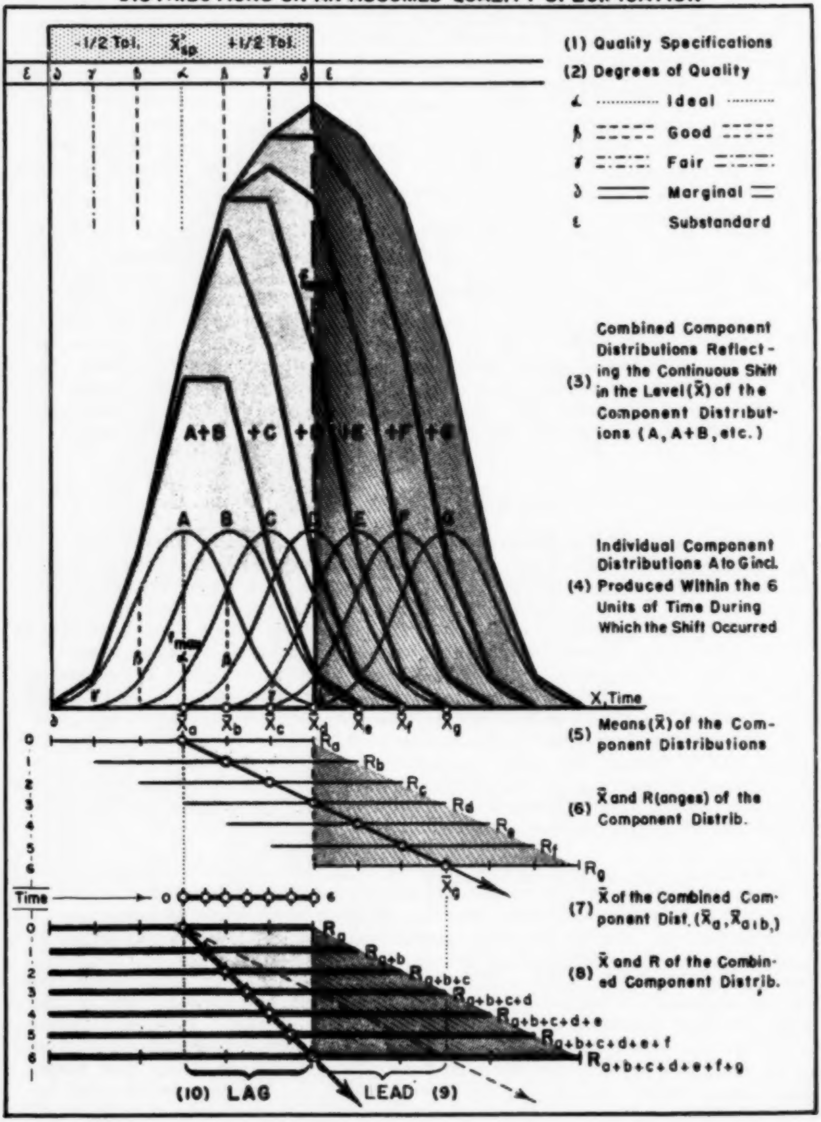
Since we are interested in the statistical method suitable for the detection of the conditions characterizing such changing processes rather than in definite case histories, let the seven components A to G inclusive in this figure represent the distribution of the individual values of the outside diameters as the tool wear progresses from the original setup at the average value $\bar{X}a$ to the average value $\bar{X}g$ at which point we assume the tool was changed. Or, let these distributions represent the output from seven machines of the same type operated by seven people or from the same machine by the same operator on seven occasions. Let us now concentrate on the two methods of sampling to be discussed and applied to this or similar conditions.

Rational Method of Sampling (Fig. 4, Item 6)

In the rational method of sampling, we give proper recognition to the fact that when the resulting distribution pattern such as illustrated in the figure by the combined curve of the components or sub-processes A to G inclusive is the result of the gradual changes taking place as time progresses, the only reasonable or rational method of sampling, sensitive to the detection of such progressive changes, would be to make time a useful ally in the continuous appraisal of such processes. If we assume that these changes follow a regular trend as schematically illustrated, once we know its rate, we can predict each component quality level as a function of time. However, we are now particularly interested in merely being able to detect the trend pictured, quickly and economically. The statistical properties of the normal distribution and the main inferences from them which have already been presented help us again to develop a proper method applicable to the situation.

Actually, the method is the same as that illustrated in Figs. 2 and 3 in which the averages and ranges of small samples and certain factors applied to them were utilized for the estimation of the assumedly unknown standard. For if the four grand averages in Figure 2 reflected reasonably accurately the true but unknown process average $\bar{X}'$, the averages of n values taken in the time sequence of their occurrence from the process with a definite trend should reflect that trend. While the graph of such averages will not be a perfectly straight line as illustrated in the sixth item of Fig. 4, a line fitted to them by the statistical method would in most similar cases be a straight line. Its slope would be a reflection on the speed with which the average or typical quality of the process components advances to the continuously higher (or lower) level. If the quality specifications were liberal, we would know how far we could let this trend continue, and also at what lowest

Fig. 4.
ADVERSE EFFECT OF THE CONTINUOUS SHIFT OF THE PROCESS COMPONENT DISTRIBUTIONS ON AN ASSUMED QUALITY SPECIFICATION



(or highest) setup value we could start the process to utilize to the utmost the tool life and the specification spread. The "modified" control chart for averages (or totals) and ranges (or sigmas) is based on this premise.

Now if the case illustrated reflects the tool wear (rise or fall of temperature, changing strength of solution, etc.) or a similar situation on an automatic or similarly stable machine, the machine itself remains unchanged; consequently, the pattern of the sample ranges should be fundamentally unaffected by the developing or sudden changes in the operating levels of the component distributions. In other words, the stability of those ranges should be an indication of the relatively stable variability of these moving component distributions. If, on the other hand, these ranges showed a departure from their previously established magnitude and pattern, that evidence should be interpreted as a change in the basic variability within the components and thus also within the combined process.

Obviously, since we talk about the averages of the process components ($\bar{X}_a$, $\bar{X}_b$ etc.) and the variability about them ($\pm 3\sigma_a$, $\pm 3\sigma_b$ etc.) while, at the same time, we also refer to the trend of these component means, the variability characterizing the seven components pictured must not be confused with the overall variability resulting from the combined results of these component sub-processes. In other words, since there is no definitely stated limit on the continuation of this trend, the resulting combined variability is by no means constant, unless by a premeditated action. We must therefore recognize two concepts of variability of such processes: one we may call the <u>potential variability</u> such as characterizes the process within <u>relatively short periods of operation</u>; the other the <u>overall variability</u>, in the figure represented by the combined top <u>curve A to G inclusive</u>. That this differentiation is justified and important will now be demonstrated and related to two methods of sampling.

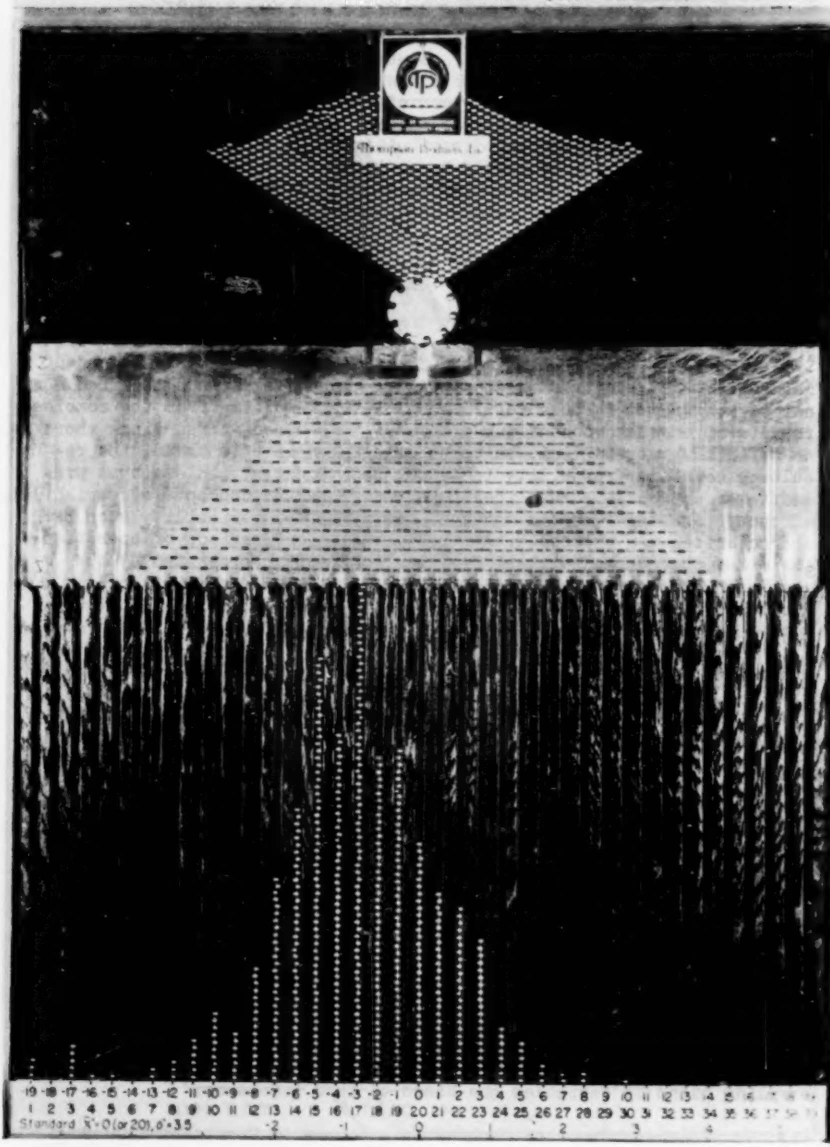
<u>Stratified Method of Sampling (Figure 4, item 8)</u>

In the stratified method, we take a large random sample from the combined process. In that case, the estimate of the average quality is quite reliable and from the mean plus minus three sigmas of that sample we also get a pretty reliable estimate of the overall variability of the whole process and its position on the scale of values. But in that case we are not able to determine whether the resulting pattern and its location is such as we might always expect of this process or whether it is subject to considerable decrease or increase in variability. For we have no picture of the developing trend of the composite quality such as indicated in the eighth item of the figure by the heavy slanting line; this results from the fact that we would have destroyed the time sequence of such values, which, in the rational method of sampling, is our ally in portraying the picture of the developing quality trend and the variability about it. In short, this method is ineffective in controlling the quality of a dynamic process. However, applied to the lots of values already produced, the method is suitable as a criterion for accepting, rejecting, or 100% inspecting such lots. Lot Plot and other similar sampling plans by variables are based on this inference.

<u>Fig. 5. Thompson Sampling Machine as a Source of Component Processes</u>

We shall now test the correctness of these inferences by creating

Fig. 5.
THOMPSON SAMPLING MACHINE (QUINCUNX)



three process components analogous to those just discussed. To be consistent, we shall represent them by 1000 values each from the Thompson sampling machine operating under three distinct conditions, one of which is illustrated in Figure 5. We shall refer to them as the process components A, B, and C. Later, we shall combine them into a combined process D representative of the condition portrayed in Figure 4 by the combination of the seven process components. We shall apply to those component and combined distributions the two methods of sampling just considered.

Fig. 6. $\bar{X}$, R Charts of m Rational and Stratified Subgroups of n Values from Three Component Processes.

In this figure, the three process components may be thought to represent dimensional values from a three spindle machine in which each spindle operates on a considerably different level, as the output of three operators working under three distinct setups, as three lots from three different suppliers, etc.

In each case, the 1000 values were recorded in 40 rows of 25 values. By calculating the averages and ranges of these rows we secure 40 averages and ranges of rational samples in which the time sequence of the occurrence of each 25 values is preserved. By calculating the averages and ranges in the column manner, we secure 25 averages and ranges of stratified samples of 40 values in which their time sequence is practically obliterated. The two sets of averages and ranges are charted in the left and right portions of the figure. The three sources are identified by the respective letters A, B, C and separate pictorial symbols.

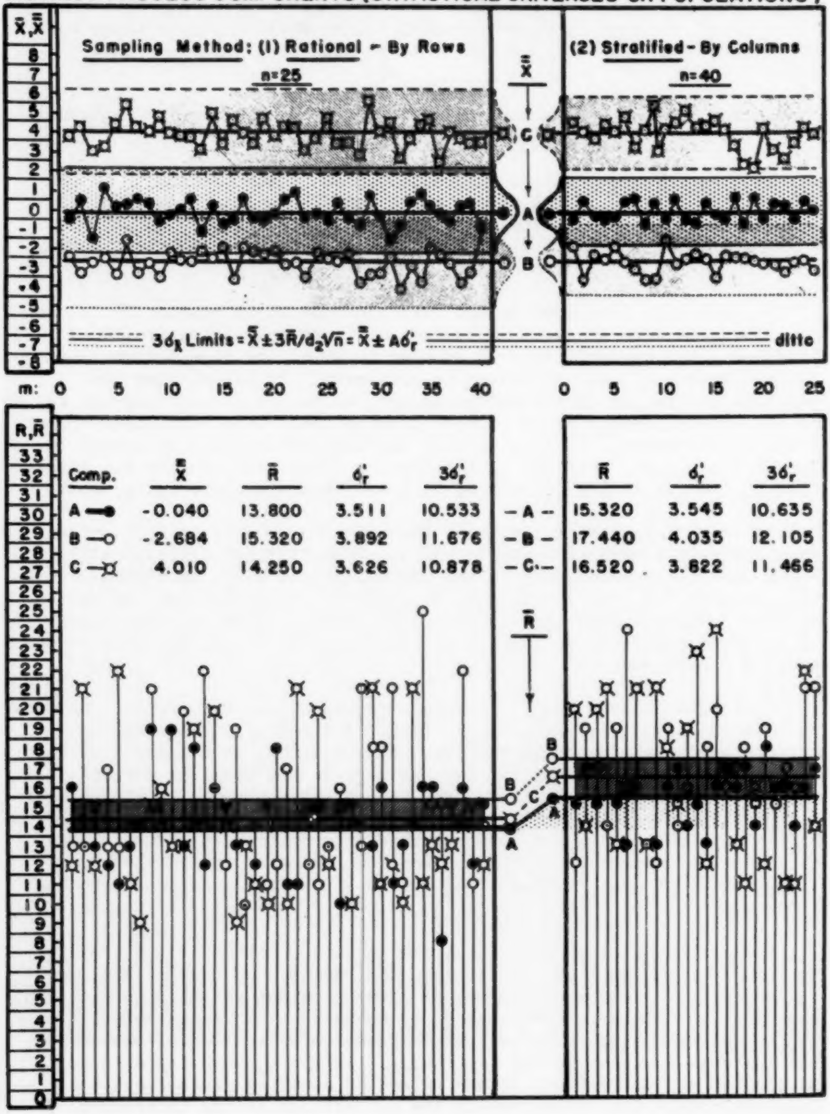
It is believed the analysis of these control charts is self-evident. They indicate a perfect control, but at three distinct levels. All these averages are well within the control limits based on the relationship between the sample size (n), the average range ($\bar{R}$), and the factor d_2 such as was already used in Figure 2, when the averages and ranges of the four smaller sample sizes selected (n = 2, 5, 10, 15) were processed in the same manner.

Note particularly that the average ranges of these components or subprocess are essentially identical. This fact confirms the previous inference that the trend in the level of operation, or a sudden shift, such as now represented by raising the legs of the sampling machine, need not influence the basic variability of these components if the rest of the process remains fundamentally unchanged. The average ranges of the stratified samples are larger than of the rational samples chiefly on account of their larger size, a fact supported by the comparable estimates of the natural limits of variability shown numerically in the figure.

Note that the pattern of these averages is practically the same for both methods of sampling. To explain that point, let us view the three sub-processes as three cakes made of three distinct doughs from different ingredients. Since in both methods of sampling illustrated we still sample only one cake at a time, it is practically immaterial whether we cut the cake horizontally or vertically. In short, when there is no problem of stratification, either method of sampling gives about the same estimate. Let us now verify the correctness of the inferences concerning the combined process (D = A+B+C) and the two methods of sampling applied.

Fig. 6.

$\bar{X}$,R CHARTS OF m RATIONAL & STRATIFIED SUBGROUPS OF n VALUES X FROM 3 DISTINCT PROCESS COMPONENTS (STATISTICAL UNIVERSES OR POPULATIONS)



Comp.	$\bar{\bar{X}}$	$\bar{R}$	σ'_r	$3\sigma'_r$		$\bar{R}$	σ'_r	$3\sigma'_r$
A	-0.040	13.800	3.511	10.533	- A -	15.320	3.545	10.635
B	-2.684	15.320	3.892	11.676	- B -	17.440	4.035	12.105
C	4.010	14.250	3.626	10.878	- C -	16.520	3.822	11.466

Fig. 7. Comparison of $\bar{X}$, R Values Based on Rational and Stratified Sampling of a Compound Process.

The process D pictured in this figure is composed of the alternating layers of 25 values from the three component distributions just analyzed, in the sequence A, B, C, A, B, C, A, B, C, etc. In the left portion of the figure are the averages and ranges of the respective 40 quality layers or strata, so that each average and range still pertains to a single source or stratum as shown by the same three symbols used in the previous figure. In the right portion of the figure, the same 1000 values are sampled in the stratified manner, in which the identity of the three component sources is obliterated. This is designated by assigning a new pictorial symbol to their averages and ranges.

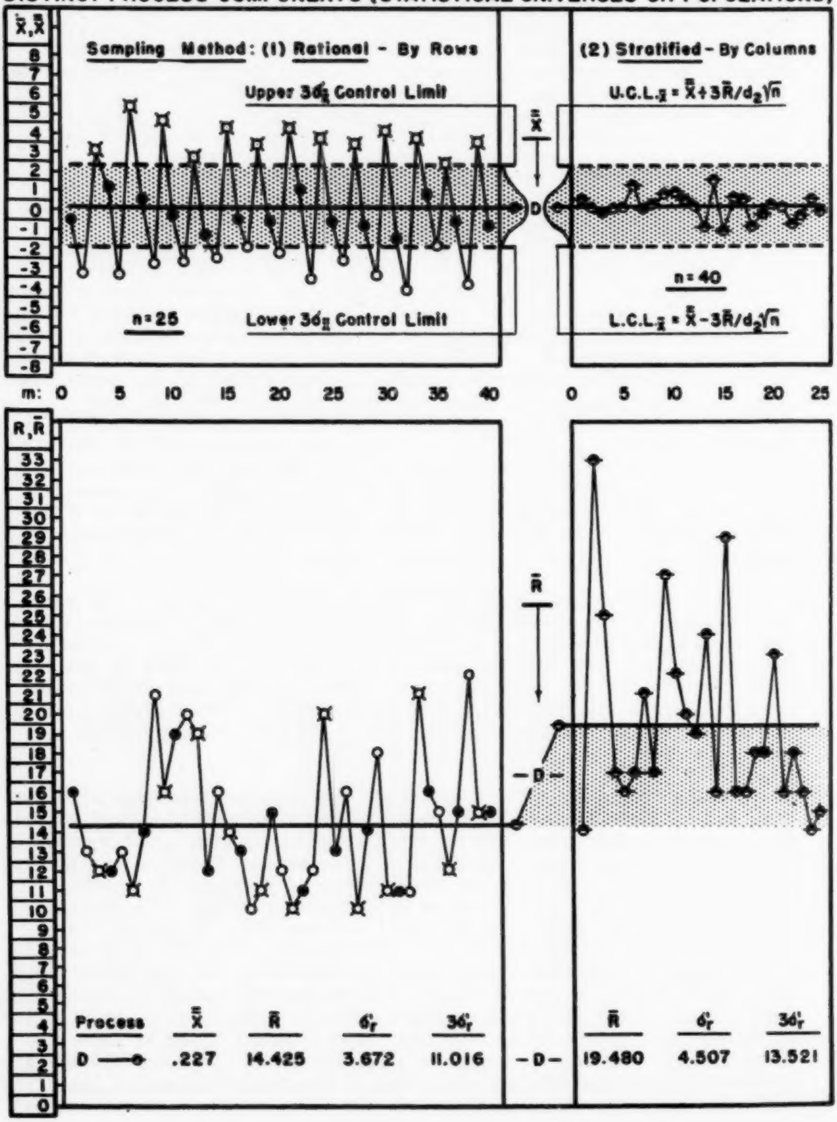
Note that two thirds of these averages of the rational samples are out of control in respect to their three sigma limits based on the average range which reflects the variability of these component distributions at the three levels. Fortunately, since in this method of sampling we did not destroy the identity of the three component sources, we can readily determine the assignable cause or causes of these out of control conditions. Plainly, all out of control points are the averages of the components B and C, none of the component A. The corrective action is therefore self-evident. The average ranges of the three component distributions being about the same, if all three components operated on the desirable level A in respect to an assumed quality specification $\bar{X}'_{sp} \pm 1/2$ Tol. $= 0 \pm 10.5$, no averages would be out of control and the overall variability of the combined process would be about the same as the variability of the single components; consequently, the assumed quality specification, 0 ± 10.5, would be easily met. In short, this method is sensitive to the detection of the situation pictured.

Let us now view the result of the stratified sampling of the same 1000 values. Not a single average of these 40 values is out of control. The averages vary so little that they are within the $\bar{\bar{X}} \pm 1\ \sigma_{\bar{X}}$ of their control limits. The picture looks very promising, in spite of the fact that the situation is quite to the contrary. We are interested in the reason for this lack of sensitivity to reveal the hidden assignable causes of the increased variation. The chart for ranges provides a partial answer.

Although those ranges are expected to be larger than those of the rational samples of only 25 values, they are very much larger than would be expected purely on account of the larger sample size of 40 values. The reason is that these stratified samples reflect both the variability within each stratum as well as among the three averages or central values of those strata. To return to the cake analogy, this combined process is a layer cake. If we cut it horizontally, we sample the variability of only one layer at a time whereas if we cut it vertically, we cut into 40 layers of three kinds; consequently, the resulting variability, reflected in these ranges, represents the variability resulting from chance causes of variation characterizing each layer as well as the extra variability resulting from the three distinct types of layers. Consequently, the average range as a reflection of that variability is now much larger than the average range of the rational samples that reflected only the variability within the three strata, but not between them.

Fig. 7.

$\bar{X}$,R CHARTS OF m RATIONAL & STRATIFIED SUBGROUPS OF n VALUES X FROM 3 DISTINCT PROCESS COMPONENTS (STATISTICAL UNIVERSES OR POPULATIONS)



It is to be noted that the process average is representative of the resulting quality and we have no quarrel with it. Rather, we have a quarrel with the method of sampling which did not furnish us with any key as to what is to be done with the process. To be sure, we know that its variability ($\pm 3\,\sigma'_r = \pm 3\,\bar{R}_D / d_2 = \pm 13.521$), if estimated from the average range ($\bar{R}_D = 19.48$) is excessive in respect to the assumed quality specification (0 ± 10.5); but we cannot put a finger on the trouble. These ranges show good statistical control, for they are based on 25 samples of 40 values made homogenous through stratification which also causes the mixing of the chance and assignable (3 levels) causes of variation creating a new system of chance causes that is reflected in the state of control of these ranges.

Fig. 8. Comparison of Potential and Actual Variability of Complex Processes Related to the Method of Sampling.

In this figure, the estimated spread of the respective component and combined distributions is shown schematically for both rational and stratified methods of sampling. Since the figure is believed self-explanatory, attention is directed only to the two top details in which the process D is shown separately for each method of sampling. Note the difference between the two pictures portraying the width of the process average $\pm 3\,\sigma$ spread about it as compared with the spread of the component distributions A, B, C, based on both methods of sampling. They vary considerably.

Since in this case the assumed quality specification 0 ± 10.5 is quite liberal, the difference between the estimates of quality values produced within it on the basis of the two methods of sampling is not large. However, the case differs when the specification is narrow. This difference stems from the statistical property of the normal distribution which is characterized by very heavy concentration in the middle and very rare occurrence of values at the ends. The next figure illustrates that fact.

Fig. 9. Relation of Quality Specification to Estimates of Variability of Compound Process Related to the Method of Sampling.

In this figure, the previous analysis is presented in the form of bar diagrams. The areas of these vertical rectangles represent the percentages of the quality values of the same three components A, B, C and their combination D based on the rational and stratified methods of sampling within two assumed quality specification; one, 0 ± 10.5, charted in the left portion of the figure, and the other, 0 ± 5, charted in the right portion of the figure. Note that in the latter case the difference between the two percentages is considerable only in the case of the process D (cake analogy).

It is hoped this discussion of two methods of sampling has demonstrated that it is very important to select a proper method of sampling if certain results are expected from it. If we merely wish to estimate the average value of the characteristic and the variability about it, the stratified method of sampling, which characterizes lot by lot acceptance sampling of the static results, is a workable method. Since in such cases the values are intermingled, no other method is practically available. On the other hand, in the sampling of a continuous dynamic process, when we still have a control over it, the only method which lends itself to a quick detection of the changing trends or sudden

Fig. 8.

$\bar{\bar{X}} \pm 3\sigma'_r$ OF COMPONENT AND COMBINED DISTRIBUTIONS ESTIMATED FROM $\bar{\bar{X}}$ AND $\bar{R}$ OF 40 RATIONAL AND 25 STRATIFIED SAMPLES OF 25 AND 40 VALUES X

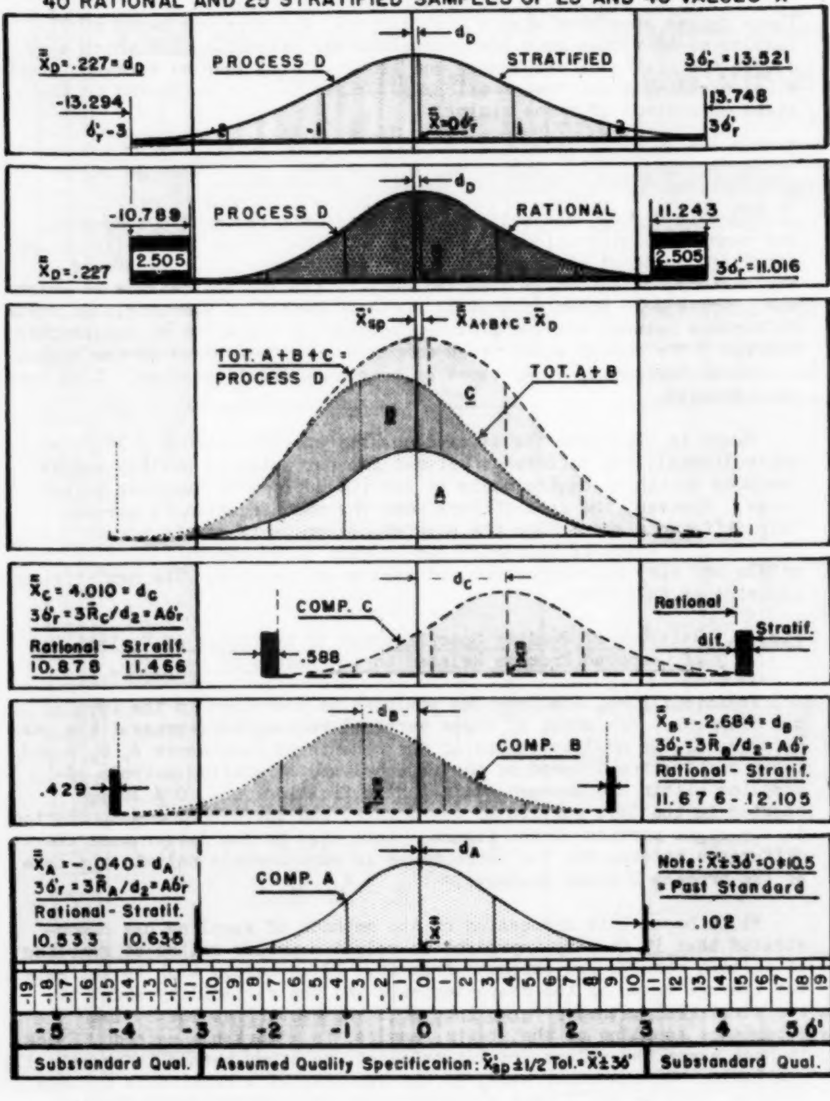
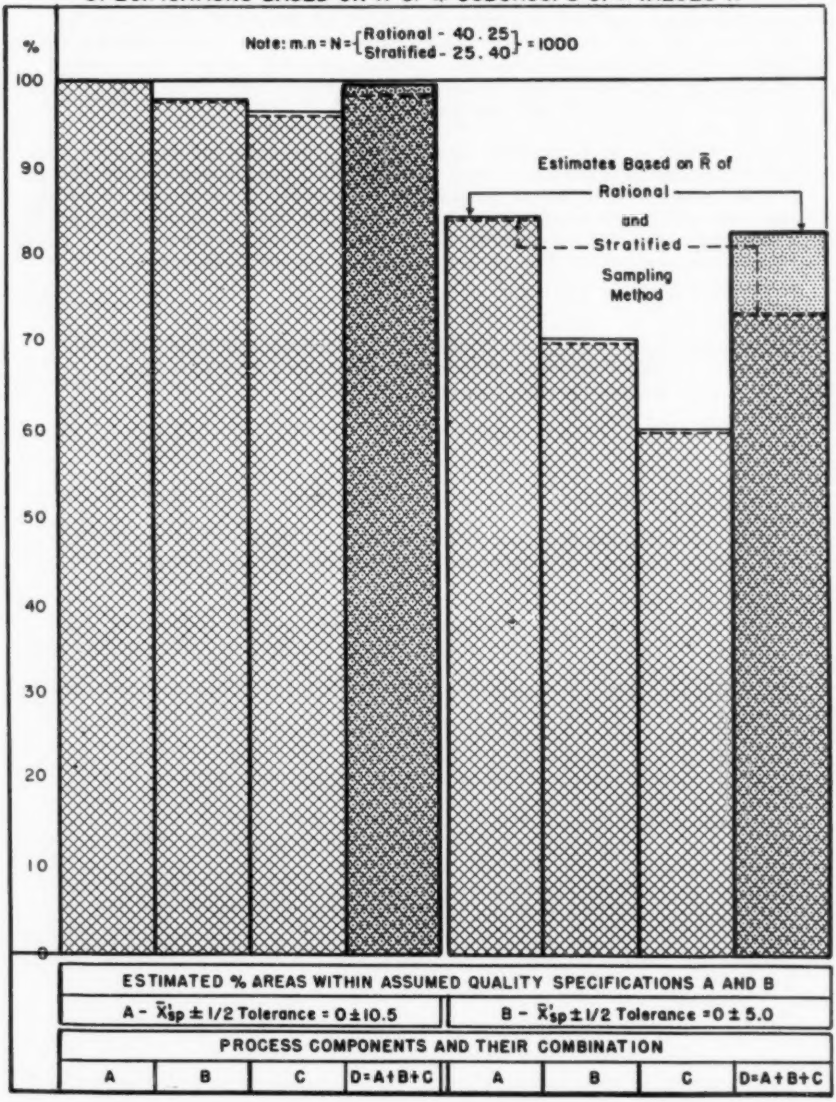


Fig. 9.

COMPARISON OF % AREAS UNDER NORMAL CURVE WITHIN TWO ASSUMED QUALITY SPECIFICATIONS BASED ON $\bar{R}$ OF m SUBGROUPS OF n VALUES X



changes of quality is the rational method of sampling, especially of the complex multiple stage processes.

Although the examples discussed are based on the data from a teaching device, the situation illustrated is anything but academical. Stratification permeates practically any activity, whether industrial or otherwise. And while a certain amount of stratification must be accepted as unavoidable, not taking it into serious consideration in operations such as mentioned is courting trouble. For if the method of sampling is carelessly chosen, the estimates resulting from it give an erroneous appraisal of the process.

Production of consistent quality requires constant vigilance over the behaviour of each contributing component. And since in stratification we lose track of those components, that method is unable to point out when the process should be left alone and when it should be investigated and corrected. Whenever it is likely to cause mischief, it should be guarded against in the manner suggested by the illustrations presented.

Reference: J. V. Strela: " STATISTICAL EVALUATION OF THE RATIONAL AND STRATIFIED METHODS OF SAMPLING", The Tool Engineer, August through December, 1952.

CHART CONTROL WITHOUT CHARTS - SIMPLE, EFFECTIVE
J. & L. QUALITY PRE-CONTROL

Dorian Shainin
Industrial Consultant
Rath & Strong, Inc.

ACKNOWLEDGMENT

The research work for PRE-Control was undertaken by the statistical engineering group* of Rath & Strong, Inc. It was sponsored by and conducted at the Jones & Lamson Machine Company of Springfield, Vermont.

Jones & Lamson manufactures machine tools and an inspection device called the Optical Comparator. Much credit for the development of PRE-Control must go to the versatile J. & L. Comparator, and to the company whose interest in production equipment naturally led to this research work in quality control.

SPECIFICATIONS

PRE-Control was designed to be a general purpose plan to replace diverse special purpose plans for controlling quality while production proceeds. It had to meet these requirements:

1. Protect against unwanted shifts in process position or centering.

2. Protect against unwanted increases in process spread.

3. Be able, for long production runs, to guarantee that the percentage of defective product produced not exceed specified values.

4. Serve, for short production runs, as a set-up plan, starting with the first piece produced.

5. Automatically adjust inspection frequency to maintain economy of control.

6. Require no paper work but permit simple records for review by supervisors and engineers, or for quality guarantees.

7. Require no measurements of product on a continuous scale; permit use of "go-not go" gaging.

8. Work from specification tolerances rather than require accumulation of data for computation of control limits.

9. Quickly identify specification tolerances which are too tight (or quite loose) compared to the "natural" process tolerance; be able to provide an estimate of the natural tolerance or process capability.

---

* C. W. Carter, W. R. Purcell, F. E. Satterthwaite, and D. Shainin

10. Allow production the use of the full specification tolerance.

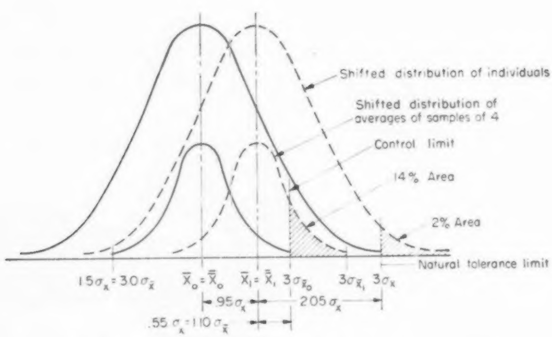
11. Be capable of being taught to operators in a short time.

12. Be competitive in efficiency with alternate plans, yet with smaller administrative and quality costs.

REFERENCES TO OTHER WORK IN THIS FIELD

The mainstay of statistical quality control plans has been the chart for averages and ranges, called by some $\bar{X}$ (X-bar) and R, and by others the Shewhart control chart by variables. Its record of effectiveness is unquestioned. In these charts control limits, as calculated from 30 to 125 preliminary measurements, define the limits of probable chance variations to be expected when there are no process disturbances. An average or a range beyond these control limits calls for action. The subgrouping of measurements in average and range charts gives improved ability to detect small process shifts. Thus Figure 1 shows how a shift, which produces 2 per cent of the work beyond a 3-sigma tolerance, will be detected about once for every 7 subgroups (14 per cent of the time) of 4 measurements; whereas 50 individual readings are required, on the average. to obtain a single individual measurement outside of tolerance. Thus subgrouping does with 28 items of data the same job that requires 50 individual readings.

Fig. 1



σ_x is the standard deviation of the curve of individual measurements.

$\sigma_{\bar{x}}$ is the standard deviation of the curve of averages of the samples

$\bar{\bar{X}}_0$ is the average of the original averages.

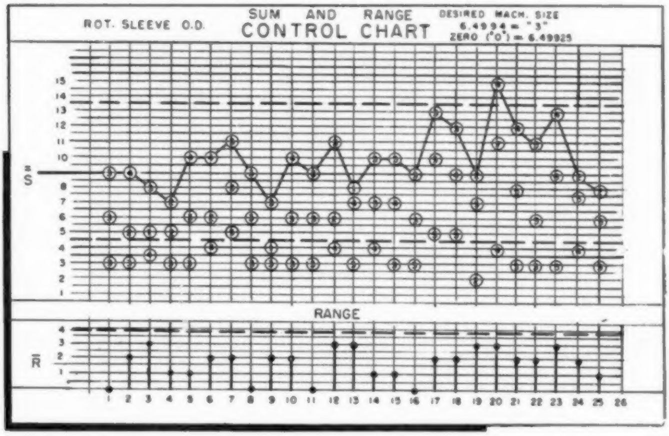
$\bar{\bar{X}}_1$ is the average of the shifted averages.

Many concerns feel that this improved sensitivity justifies the work of computing averages and ranges on work sheets and transferring the results to the control chart. In recent years, however, many alternate schemes have been introduced that require less work than does the method of average and range control charts.

I personally had been using a modification, called the sum and range chart (1), in many plants. See Figure 2. It appeals because of

the elimination of the work sheet. because of the immediate entering of measurements. and because of the absence of division of each subgroup sum by the sample size to get the average. This chart has the identical good statistical characteristics as the average and range chart it replaces.

Fig. 2



Ferrell (2) has proposed a plan using the mid-range and range (the mean and difference of the extremes). Median charts have advantages over average charts for life testing (3) and some non-normal distributions.

The use of attribute (go-not go) measurements for control is very attractive. Carlson (4) uses standard lot-by-lot acceptance plans, spreading the required sample over the production period. A great many of these plans use special limits and have been presented under such names as (5) control gaging, compressed limit gaging, narrow limit gaging, limit gaging, and <u>n p</u> chart control. The special limits, computed from a preliminary estimate of process dispersion (based on special measurements) are set within the tolerance limits. Lack of control is indicated when the number of pieces outside these special limits exceeds the number allowable under the specific control plan.

PRE-Control is related to these plans. But control limits are based on the specification tolerance, eliminating the need of preliminary estimates of the natural tolerance. Also it uses sequentialization to reduce the amount of inspection and it adjusts inspection frequency to give Average Quality Limit guarantees.

ELEMENTS OF CONTROL PLANS FOR OPTIMIZING THEIR DESIGN

Consider perhaps the simplest control plan. An inspection is made. If the result falls within the specified tolerance the process continues without adjustment; if it is outside, the process is corrected. This plan has the three elements required in any control plan: inspection checks, a control function, and control limits.

1. The checks supply the information on which the control decision is based. In this example the decision is based on a single result. Other plans use more than one. The more checks made, the more information available, and the greater is the potential for a good decision.

2. A control function summarizes information in the measurements. In this example the control function is the check result itself. But it may be the average of several checks, or the range of a group of measurements, or the number that fall beyond a control limit. or any other appropriate function.

3. Control limits convert the control function to a control decision. Here the control limits are the specified tolerance limits. The decision to correct or not to correct the process depends upon whether the control function falls outside or within the control limits. In general the control limits do not coincide with the tolerance limits but are chosen to give a good balance between benefits and cost.

For the steps that lead to a better design for a control plan, a look should be taken at costs and benefits.

The costs of control naturally divide into two groups, operating costs and inherent statistical costs. The former include the costs of inspection checks, of recording them, of analyzing and interpreting them, and of taking the indicated actions to correct the process. The statistical costs arise when the control plan occasionally indicates there has been a change in the process when in fact no such change has occurred. The action from such a "false" control indication causes the process to operate more poorly than necessary. But the plan should soon detect this poor operation and indicate a corrective action to counteract the original action. Logically the costs of such actions from false control indications are called "hunting" costs. They are statistical costs inherent to some degree in all control plans.

Benefits also can be conveniently listed in two categories, but this time both of a statistical nature: the sensitivity of the control plan and the quality level it can guarantee. Of course there are also benefits of a non-statistical nature, such as the reduction of scrap and reoperation, which are the prime incentive for having a control plan. But these will have little effect on the specific design of the control plan.

Sensitivity is the basic measure of control effectiveness. If a process would always operate in the best possible manner, control would be unnecessary. But such processes are exceptional. Most practical processes react to spontaneous changes, which act at unknown times and with effects of unknown character and magnitude. Control checks aid in the detection of such changes, but not instantaneously. A <u>sensitivity measure</u> of the control plan is related to the average of the time lags between the changes and their detection.

The most fundamental sensitivity measure is the damage (evaluated in dollars) caused by a process change before its detection. Since this paper is not considering specific applications, this damage is approximated by estimating the average number of defective pieces produced.

Most published process control plans do not give average quality limit guarantees; that is, guarantees that the fraction defective will not exceed specified values. PRE-Control plans yield both Average Production Quality Limit (APQL) and Average Outgoing Quality Limit (AOQL) values, APQL referring to the maximum fraction of defective parts produced by the process, and AOQL referring to the maximum fraction outgoing after removal of defectives found by the inspections called for in the control plan.

In summary, the aim in control plan design is to obtain control sensitivity and preferably an average quality limit guarantee (APQL or AOQL) with a minimum of hunting and administration cost. To accomplish these purposes, the following factors can be varied:

1. Measurement precision
2. Inspection frequency
3. Control limits
4. The control function
5. Measurement timing

Control plan design can in only a minor way substitute for measurement precision, which improves as the square root of the number of measurements. While it may be practical, for instance, to increase the number of measurements fourfold to double precision, it is usually completely impractical to increase measurements 100 times to get a tenfold increase in precision. Substantial increases in precision require improvement of the measurement method (e.g., use of electro-limit gages instead of micrometers).

Control of inspection frequency is the convenient method for quality guarantee. This is done by keeping constant the average number of inspections per out-of-control indication. Then an increase in process shift frequency (with consequent proportional increase in the number of defectives) is exactly counteracted by an increase in inspection frequency, which proportionally reduces the time required to detect the process shift (and therefore the number of defectives produced before the detection).

Most processes, of course, have process shifts of varying sizes, and the fraction defective produced depends on these process shift sizes. In Table 1 are illustrated the situations when 25 inspections are made per out-of-control indication, tolerances are at ±3.2 sigma, and control limits for individual values are located as shown. Now the important fact to notice is that there is a process shift size that produces a maximum fraction defective as shown by the boxed results in Table 1. Thus, with control limits at ±3.0 sigma, a 2.0-sigma shift produces 2.6% defective product (on the average) and this is larger than the fraction defective produced by any other process shift size.

The average production quality limit (APQL) guarantee for the plan in Table 1 with control limits at 3.0 sigma is therefore 2.6% defective. In a real process the true size of the process shift is unknown, but whatever that unknown size, it will produce a fraction defective not more than 2.6%

Table 1 also illustrates how control limit location affects the sensitivity of a control plan to small process shifts. Narrowing the control limits from ±3.2 sigma to ±3.0 sigma reduces the maximum fraction defective (APQL) from 3.9% (for small process shifts) to 2.6% (realized at

2.0-sigma shifts). Further narrowing the control limits to ±2.0 sigma greatly reduces the fraction defective for small shifts, and the maximum (for large shifts) is 2.0%. But note that this reduction in fraction defective for small shifts is obtained only at the expense of increased "hunting" (which increases very rapidly if the limits are narrowed further, being 317 for ±1.0 sigma limits). Also note that control limit location has no effect at all on sensitivity to very large process shifts.

Table 1

Fraction Defective Produced When There Are 25 Inspections Per Out-of-Control Indication

Size of Process Shift	Control Limits for Individual Measurements at ±3.2 sigma	±3.0 sigma	±2.0 sigma
0.0 sigma	0.1%	0.1%	0.1%
1.0 "	3.9%	2.4%	0.5%
2.0 "	3.7%	2.6%	1.0%
3.0 "	3.2%	2.5%	1.3%
4.0 "	2.4%	2.2%	1.7%
5.0 "	2.1%	2.1%	1.9%
6.0 " and over	2.0%	2.0%	2.0%
Hunting (False out-of-control indications per 1,000 inspections)	1.4	2.5	45.5

Control plan characteristics depend on the control function used and the timing of the measurements. There are five possible basic variations in these factors which will now be described in conjunction with the illustrative numerical examples shown in Table 2. (These examples have been made equivalent by adjusting the control limits to give equal hunting and by adjusting inspection frequency to give equal numbers of pieces inspected.)

1. Individual measurement plan.

2. Multiple measurement control functions: Plan (2) illustrates how increasing the number of measurements included in the control function increases control sensitivity to small process shifts (up to 3.0 sigma) at the expense of reduced sensitivity to medium large process shifts (3.0 sigma to 9.0 sigma).

3. Grouping measurements: Plan (3) illustrates how grouping measurements (for example, four measurements each hour instead of one measurement each 15 minutes) increases sensitivity to small shifts at the expense of much reduced sensitivity to large shifts.

4. Multiple control functions: Skillful use of two (or more) control functions can often realize the advantages of each. Plan (4) combines plans (1) and (2). Lack of control is indicated if the last measurement is outside of ±3.81 sigma or if the average of the last four is outside of ±1.51 sigma. This plan is practically as good as (2) for small process shifts and as good as (1) for large shifts. It is only slightly poorer than (1) for medium large shifts.

5. Sequential plans: Plan (5) is a sequential plan which calls for an immediate check sample of four additional measurements whenever a single measurement is outside of ±2.24 sigma contol limits. In no case is this plan more than 10% poorer than the best of the other plans and it is up to 45% better for medium (3.0 sigma) process shifts.

Table 2

Size of Process Shift	Average Fraction Defective Produced: Individuals (1)	Multiple Measurement (2)	Grouped Measurements (3)	Multiple Control Functions (4)	Sequential (5)
0.0 sigma	0.1%	0.1%	0.1%	0.1%	0.1%
1.0 "	2.4%	0.8%	0.6%	0.8%	0.7%
2.0 "	2.6%	1.4%	1.1%	1.3%	1.0%
3.0 "	2.5%	3.1%	3.3%	3.0%	1.5%
4.0 "	2.2%	4.6%	6.3%	3.0%	1.8%
5.0 "	2.1%	4.6%	7.7%	2.4%	2.1%
6.0 "	2.0%	4.0%	8.0%	2.1%	2.2%
9.0 "	2.0%	2.2%	8.0%	2.0%	2.2%
12.0 "	2.0%	2.0%	8.0%	2.0%	2.2%
Measurement Spacing	Uniform	Uniform	Groups of 4	Uniform	(1) + (3)
Control Functions	Single Measurement	Average of last 4	Average of group	(1) + (2)	(1) + (3)
Control Limits	$\pm 3.0\sigma$	$\pm 1.5\sigma$	$\pm 1.27\sigma$	$\pm 3.81\sigma$ $\pm 1.51\sigma$	$\pm 2.24\sigma$ $\pm 0.59\sigma$
Inspections per Out-of-Control Indication	25	25	100	25	23

Quality control literature has put great stress on grouped measurement plans (particularly the average, X-bar, and the range, R, control chart plans). When applied to processes where small process shifts are important (particularly if specifications are tight or measuring precision is poor), their success has often been excellent. But in many industrial situations, small process shifts are unimportant. This appears to be the chief reason why successful applications of X-bar and R charts have been spotty. Conventional X-bar and R charts give poor and uneconomical control of large process shifts.*

The most efficient possible timing of control measurements is to make no measurements on the process except immediately following a process shift. This is possible (and almost always done) for those processes where all (or most) process shifts are known to the operator without making measurements (breakage, tool changes, material changes, set-up changes, etc.). But routine control plans are required because a process is subject to unknown spontaneous shifts occurring at random times. In this situation the most efficient control plan (as indicated in Table 2) calls for periodic single measurements with additional check measurements whenever a single measurement hints that a process shift has occurred. This tends to concentrate the check measurements at times when a process shift has actually occurred. Therefore, the check measurements may supply substantially more control information than an equal number of measurements taken periodically.

Control plans utilizing this check measurement principle are called sequential control plans.

Table 2 shows that it is possible to use the increased information in the check measurements to design sequential plans with improved sensitivity to small process shifts and with only a negligible loss in sensitivity to large process shifts. This avoids the serious disadvantage of multiple and grouped measurement plans for general purpose applications.

Table 2 considered the fraction defective produced under the control plan. Further reductions can be obtained by retroactive inspection which involves 100% inspection of all pieces produced since the last previous inspection whenever an out-of-control indication is obtained. Removal of any defectives found in this 100% retroactive inspection gives, for the larger process shifts, an average outgoing fraction defective that is substantially better than the average produced fraction defective as shown in the example of Table 3.

While the numerical examples considered in the foregoing discussion considered only shifts in the process position (or center), the same principles apply for control of increases in process spread.

---

* Assuming inspection costs are proportional to the number of pieces inspected. If inspection costs are proportional to the number of groups of successive pieces inspected, grouped measurement plans may be desirable to increase measurement precision.

Table 3

Size of Process Shift	Average Fraction Defective Produced	Outgoing
0.0 sigma	0.1%	0.1%
1.0 "	2.4%	2.3%
2.0 "	2.6%	2.2%
3.0 "	2.5%	1.3%
4.0 "	2.2%	0.4%
5.0 "	2.1%	0.0%
6.0 " and over	2.0%	0.0%

PRE-CONTROL

PRE-Control is the first control plan proposed that takes advantage of all the foregoing principles. It is a sequential plan to obtain high sensitivity with minimum inspection. It adjusts inspection frequency to give average produced quality limit guarantees. If necessary, variations may be used to give even tighter average outgoing quality limit guarantees through retroactive inspection. Its control limits have been chosen to give maximum sensitivity to small shifts without objectionable hunting. It is an attribute plan which simplifies gaging (go-not go), arithmetic (counting only), instruction, and administration.

The operating details of PRE-Control are diagrammed in Figure 3. The P-C lines are half the tolerance width. Twelve successive pieces within the P-C lines indicate the process is sufficiently well adjusted to allow sampling inspection. On sampling inspection, a single piece outside P-C calls for an immediate check inspection and if this is also out, the process should be corrected. Adjusting sampling inspection frequency to give about 25 inspections per out-of-control indication generally keeps the fraction defective produced below 1% and guarantees an average produced quality limit (APQL) of 2% (normal distribution, conservatively 3% to allow for some non-normality).

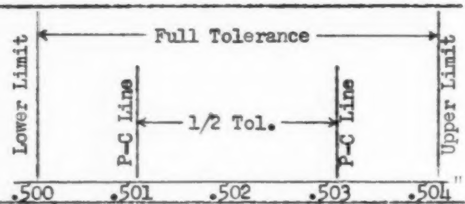
If the specification tolerance is too tight for the natural tolerance of the process, this will be quickly apparent because successive measurements on opposite sides of the P-C line band will occur repeatedly during the attempt to obtain 12 successive pieces within P-C lines.

The basic PRE-Control Plan A (Figure 3) is a satisfactory general purpose control plan for most processes. Three variations have also been selected to cover important but less common situations:

Variation B. Increased sampling inspection, a tighter control function, and retroactive inspection give an average outgoing quality limit (AOQL) of 1/3 of 1% for those critical specifications requiring

Figure 3. STEPS IN PRE-CONTROL

Have a pair of guide lines centrally drawn inside the tolerance limits, but just half as far apart. These guide lines are called P-C for Pre-Control. They do not cut the tolerance - they only provide information.



Inspect first few pieces from process. Count how many consecutive results come between the P-C Lines.

Whenever 12 results in a row all fall between P-C Lines, go to sampling control.

Sample at whatever frequency gives about 25 inspections on the average for each correction (2 in a row beyond P-C) *

Whenever a result is outside a P-C Line #, inspect the very next piece and: -

If it is between the P-C Lines, disregard the piece that was beyond the P-C Line.

If it is also beyond the same P-C Line, the pattern is out of position. ADJUST SETTING

If it is beyond the opposite P-C Line, the pattern is wider than the tolerance. RESHARPEN OR CHANGE TOOL; WORK CLAMP IS NOT HOLDING UNIFORMLY, etc.

\# If at any time a result is beyond a tolerance limit, it may be desirable to adjust the process at once.

* In the long run this practice will keep the number of parts which are slightly beyond tolerance at probably less than one per cent and definitely less than three per cent. Of course, if the process is extremely stable, running for long periods without requiring adjustment, a decision may be in order to inspect twice a day, perhaps, or once each shift, or at some other logical interval. The process will then deliver at an even lower rate of defectiveness than with the 1-in-25 rule. Naturally, when any known change takes place - new stock, new tool, change of coolant and so on - inspect every piece until 12 in a row come between the P-C Lines.

tighter control than Plan A. Variation B differs from Plan A in the following ways:

1. Fifty inspections are required, on the average, per out-of-control indication (instead of 25)

2. If an inspected piece falls beyond a P-C line, the next 5 are checked (instead of a single check piece)

3. If any one of these five is beyond a P-C line:

   (a) Correct the process and
   (b) Check all pieces produced since the last sampling inspection, removing any pieces found outside of tolerance limits.

Variations C and D. Many processes can hold a "natural tolerance" substantially narrower than the specified tolerance. It is then safe (and often desirable) to move the P-C lines out toward the tolerance limits to allow increased room for toolwear and other small process shifts that do not produce defectives. Variations C and D do this, locating the P-C lines 1.67 sigma in from the tolerance and "safety lines" 5.0 sigma in from the tolerance. This requires determining the "six sigma" natural tolerance which may be done by any standard method. (A Capability Chart that graphically calculates the natural tolerance and the limit line locations from P-C type attribute data has been developed.)

Variation C corresponds to basic Plan A, and Variation D to tight Variation B. The only changes in operation are in the interpretation of the check measurements made when a measurement falls outside a P-C line. These check measurements (1 for C and 5 for D) must all fall between the P-C line and the corresponding safety line. If a check measurement falls between the two safety lines, this indicates an increase in process width so large that it is no longer safe to use the widened P-C lines for control.

Other Variations. While the foregoing plans are sufficient for most general purpose uses, additional modifications of control limit location and inspection frequency can be made to give operating characteristics tailored to most any possible special purpose situation.

CONTROL OF VISUAL CHARACTERISTICS

... that can't be measured on the Comparator

Experience with Comparator Quality PRE-Control will suggest a variety of applications. Because the technique has the statistical power of control by measurements without requiring readings, it can serve even when readings are impossible or impracticable. For example, where the appearance of a surface should be controlled, the technique of using a P-C Line "Standard Finish" with Plan A will control the times when the cutting edge condition, feed, speed, etc., need to be changed.

Here is a quick way to get this P-C Line Standard from a process that is currently producing acceptable finishes:

1. Take seven (or more) samples of nine consecutive pieces each from the process. If more samples are taken, keep the number of them odd.

2. Select the two pieces from each sample that represent the worst and best finishes in that sample.

3. You now have seven (or more) pairs. Arrange these in order from 1 to 7. One is that pair with the least difference in quality. Seven is that pair with the greatest difference.

4. From these seven (or more) pairs, pick the "median" pair - #4 if you used seven. This is the reason for an odd number of samples.

5. The median pair can be considered, for all practical purposes, to have a range of finish equal to one half the capability of that process to hold a "uniform" finish. This useful relation holds statistically because the number of pieces in each sample is kept at nine.

6. Again, considering an estimate that is practical and useful, use the worst finish of the two in the median range for your reference standard, or the P-C line. Two successive pieces from future production that are worse than this standard means that the process needs adjustment or that the machines need checking.

CONCLUSION

The advantages of statistical quality control have been widely demonstrated. PRE-Control has been developed and presented to bring statistical quality control closer to the goal of practical tools for everyday use on the factory floor.

REFERENCES

1. Grant, E. L. "Statistical Quality Control", 2nd edition, 1952, pg. 189, McGraw-Hill Book Co., Inc., N. Y. C.
   Shainin, D. "Quality Control Methods", Part 3, Machine Design, Vol. 24, No. 9, Sept. 1952, pgs. 172, 173, Penton Publishing Co., Cleveland 13, Ohio

2. Ferrell, E. B. "Control Charts Using Midranges and Medians", Industrial Quality Control, Vol. IX, No. 5, March, 1953, pgs. 30-34, ASQC, 70 E. 45th St., N. Y. 17, N. Y.

3. Purcell, W. R. "Saving Time in Testing Life", Electrical Engineering Vol. 68, No. 7, July 1949, pgs. 617-620, AIEE, 33 W. 39th St., N. Y. 18, N. Y.

4. Carlson, F. "An Attribute Chart for Acceptance Purposes", Industrial Quality Control, Vol. IX, No. 4, Jan. 1953, pgs. 6-8, ASQC, 70 E. 45th St., N. Y. 17, N. Y.

5. Dudding, B. P. and Jennet, W. J. "Quality Control Chart Technique when Manufacturing to a Specification", August 1944, Gryphon Press, Arlington, Va.
   "Quality Control by Limit Gaging", Production and Engineering Bulletin, Vol. 3, Oct. 1944, pgs 433-437, Great Britain

Stevens, W. L. "Control by Gaging", Journal of the Royal Statistical Society, Vol. X, No. 1, Series B, 1948, pgs. 54-108.
Grant, E. L. "Statistical Quality Control", 2nd edition, 1952, pgs. 256-258, McGraw-Hill Book Co., Inc., N. Y. C.
Mace, A. E. "The Use of Limit Gages in Process Control", Industrial Quality Control, Vol. VIII, No. 4, Jan. 1952, pgs. 24, 28-31, ASQC, 70 E. 45th St., N. Y. 17, N. Y.
Ott, E. R. and Mundel, A. B. "Narrow Limit Gaging", Quality Control Conference Papers, 1952, pgs. 321-330, ASQC, 70 E. 45th St., N. Y. 17, N. Y.
Robinson, M. H. and Heide, J. D. "The Use of Compressed Limit Gages for Process Control", Proceedings, 6th New England Quality Control Conference, pgs. 158-174, ASQC, 70 E. 45th St., N. Y. 17, N. Y.
Shainin, D. "Quality Control Methods", Part 6, Machine Design, Vol. 24, No. 12, Dec. 1952, pgs. 173-179, Penton Publishing Co., Cleveland 13, Ohio

COST OF QUALITY

W. H. Lesser
General Electric Company

Every quality control supervisor must be prepared to answer the following questions, asked annually by his General Manager, namely:

1. How much does your quality control activity cost?

2. What do I get out of it in terms of increased profit, higher inventory turnover and improved customer satisfaction?

Management will measure the costs and results of a quality control program, and compare their effect upon profit margin with that of other programs, such as: production control, standardization or industrial engineering.

An ambitious supervisor will sometimes attempt to justify his budget request by statements, such as: "Quality is difficult to measure", or "It is a good idea to spend money on quality control." Smart management will seldom accept such unfounded statements, but will listen to statements, such as: "Our cost of quality for 1953 was so many dollars", or "This vendor rating program has increased inventory turnover so many percent". These are facts that keep a General Manager interested in current quality control activities.

The following basic elements of quality cost, Chart 1, exist in any manufacturing organization.

> <u>Identified Costs,</u> as: scrap, rework, inspection, test, customer complaint expense and quality control costs, including: quality training programs, vendor rating programs, product audits, life testing and maintaining statistical analysis activities.
>
> <u>Hidden Costs,</u> as: extra cost due to poor quality planning, for example: performing additional operations on work produced by machines incapable of holding the required tolerance, or final adjust operations, necessary because of a loosely controlled previous operation. Production and shipping delays due to defective work, occasioned by inspection rejects and the time required to obtain replacements. Lost business due to a poor quality reputation. Inherent design weakness, which specifies material and manufacturing operations unnecessarily difficult or costly to keep under economical control.

The above enumerated "hidden costs" are obviously both present and difficult to obtain in most companies. However, such costs must be continually recognized and evaluated, in order to fully realize cost reductions from an integrated quality improvement program.

On Chart 2 we can compare the quality cost with other financial statistics. So many dollars of cost is equivalent to: - percent of the output of a contributing department, - percent of the expenditures for plant and equipment, - percent paid stock holders. Management is familiar with these financial facts, and full use should be made of similar comparisons in order to highlight the quality problem.

The source of quality costs is shown, Chart 3, in an extract from a typical profit and loss statement.

> Direct material, as reported in a standard cost system of accounting, readily identifies: scrap, due to poor workmanship, improper planning or defective vendor supplied material, and excess spent over standard material cost for delivery, availability or quality considerations. Quality control activities can reduce scrap and increase inventory turnover by the use of: improved vendor contacts and quality certification plans, review of specifications and methods with suppliers and adequate incoming material inspection.
>
> Direct labor contains planned standard labor, rework and excess over planned costs. A comprehensive quality control program will reduce the non-productive direct labor with the aid of effective process controls, improved job instruction, promotion campaigns and the reduction of inspection rejects. A quality minded organization is bound to reflect itself in good performance by direct labor.
>
> Overhead represents the cost of those services which do not appear directly in the finished product. Some of the following accounts will be directly affected by a successful quality improvement program.
>
> 1. Inspection cost, often reduced by the introduction of sampling plans, vendor certification plans and control charts.
> 2. Idle time and high production costs, caused by defective workmanship or weak control over manufacturing processes.
> 3. Purchasing efforts expended to obtain satisfactory replacement materials.
> 4. Material handling and stocking charges for handling defective and replacement material.
>
> Engineering expense, including customer complaint cost, can be reduced by utilizing planned experiments and pilot runs, statistical analysis of test data and tolerance review on complicated assemblies.

As shown on Chart 4, the percent inspection rejects from the assembly of an electrical product can be translated into dollar cost. The indicated annual improvement is due to an integrated program to improve product design, tools, assembly methods and incoming material condition.

The amount of direct scrap and rework labor caused by defective purchased material is easy to obtain, and is a good yard-stick to measure the effectiveness of a vendor relations program. On Chart 5 we see the amount of losses, and percent recovery by commodity, by source and by type. Often purchased material losses follow a pattern by which a small percent of the items account for a large percent of the loss. A commodity analysis will show which vendors cause most of the trouble.

Vendor supplied material in most companies falls into natural classifications, such as: affiliate or outside supplier, wharehouse or mill source, prime or subcontracted vendor. Periodic reviews are desirable in order to economically modify specifications, and to take advantage of changes in suppliers' manufacturing methods.

Note the higher percent recovery on scrap as compared to rework losses. This condition usually exists in assembly plants because of continual pressure to meet production schedules. This is where a progressive vendor relations program will save money by halting defective purchased material before shipment is made to the user.

Your key elements of quality cost can be presented to your management in a similar manner. Make full use of charts and dollars, but always bear in mind that the quality yardstick is only one of many used by management to measure overall company performance. It is only after the quality supervisor displays a knowledge of the facts upon which he will be measured, that he can expect the General Manager to whole heartedly support his quality control program.

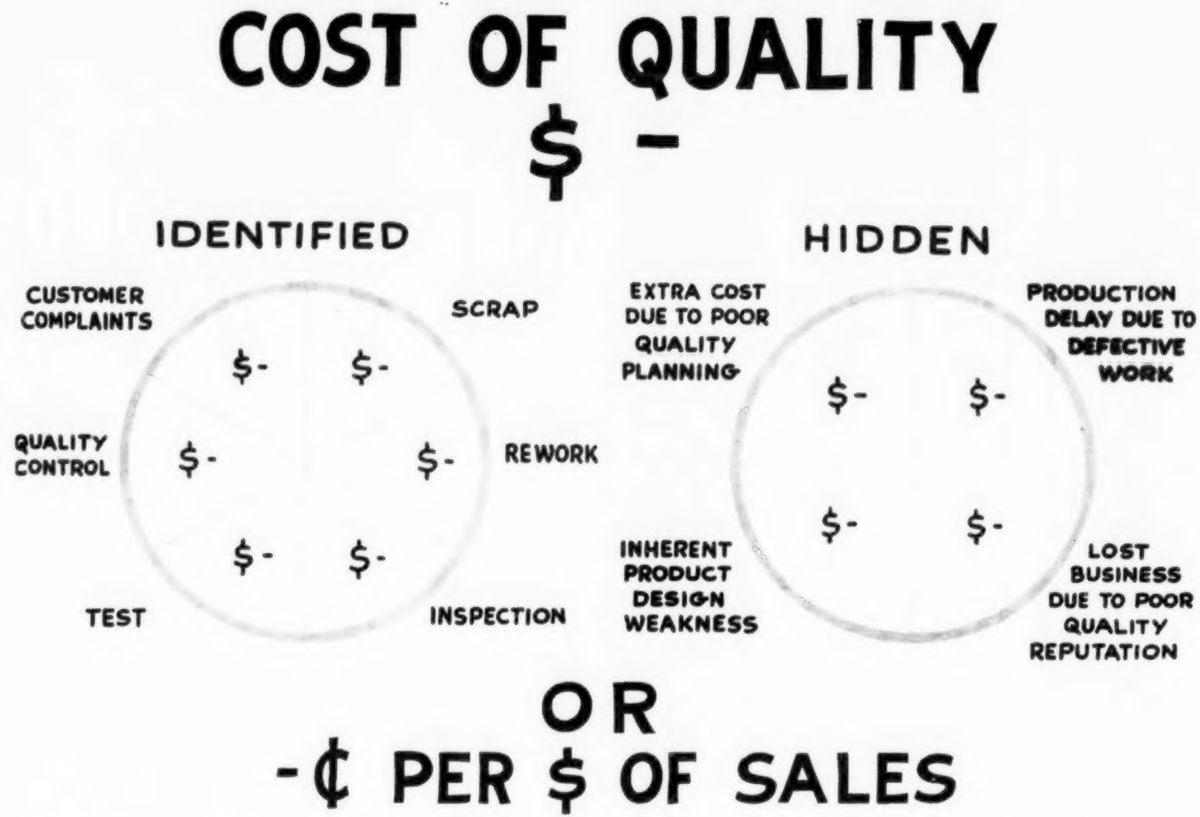

COST OF QUALITY
$ -
IDENTIFIED
CUSTOMER COMPLAINTS
SCRAP
QUALITY CONTROL
REWORK
TEST
INSPECTION
$ -
$ -
$ -
$ -
$ -
$ -
HIDDEN
EXTRA COST DUE TO POOR QUALITY PLANNING
PRODUCTION DELAY DUE TO DEFECTIVE WORK
INHERENT PRODUCT DESIGN WEAKNESS
LOST BUSINESS DUE TO POOR QUALITY REPUTATION
$ -
$ -
$ -
$ -
OR
-¢ PER $ OF SALES

CHART I

COST OF QUALITY

$ -

EQUIVALENT - TO

- % OF THE OUTPUT OF THE- DEPARTMENT

- % OF WAGES AND SALARIES PAID

- % OF THE EXPENDITURES FOR PLANT & EQUIPMENT

- % OF DIVIDENDS PAID TO STOCK HOLDERS

CHART 2

PROFIT & LOSS STATEMENT

* SOURCES OF QUALITY COSTS

NET SALES BILLED

Less: COST OF SALES

DIRECT MATERIAL	STANDARD MATERIAL EXCESS OVER STANDARD * SCRAP *
DIRECT LABOR	STANDARD LABOR REWORK * EXTRA COST * STARTING COST *
OVERHEAD	INSPECTION,* QUALITY CONTROL* TRAINING, IDLE TIME PURCHASING,* PRODUCTION *
ENGINEERING	PRODUCTION ENGR., LABOR- ATORY TESTS,* PILOT RUNS,* CUSTOMER COMPLAINTS *

NORMAL GROSS MARGIN

Less: COMMERCIAL { CONCESSIONS *to* CUSTOMERS *

ADMINISTRATIVE

ADVERTISING

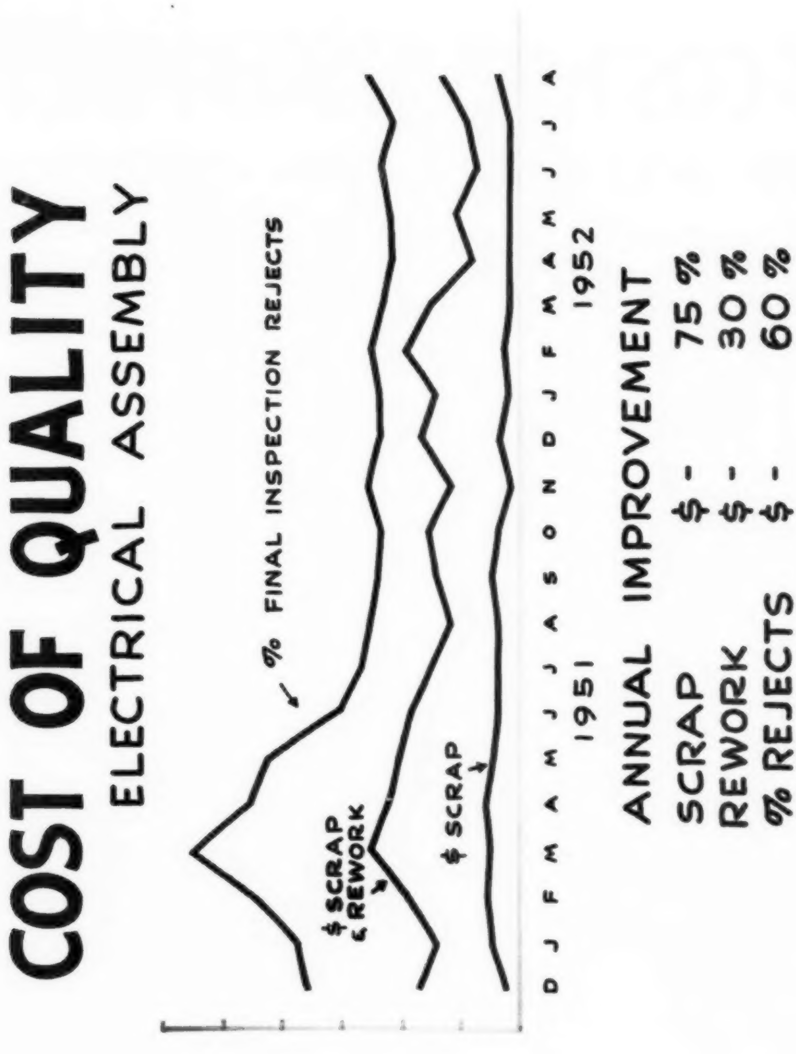
ACCOUNTING

NET INCOME BEFORE TAXES

Less: TAXES

NET PROFIT

CHART 3



COST OF QUALITY
ELECTRICAL ASSEMBLY
% FINAL INSPECTION REJECTS
$ SCRAP & REWORK
$ SCRAP
%
$
D J F M A M J J A S O N D J F M A M J J A
1951
1952
ANNUAL IMPROVEMENT
SCRAP $ - 75 %
REWORK $ - 30 %
% REJECTS $ - 60 %

CHART 4

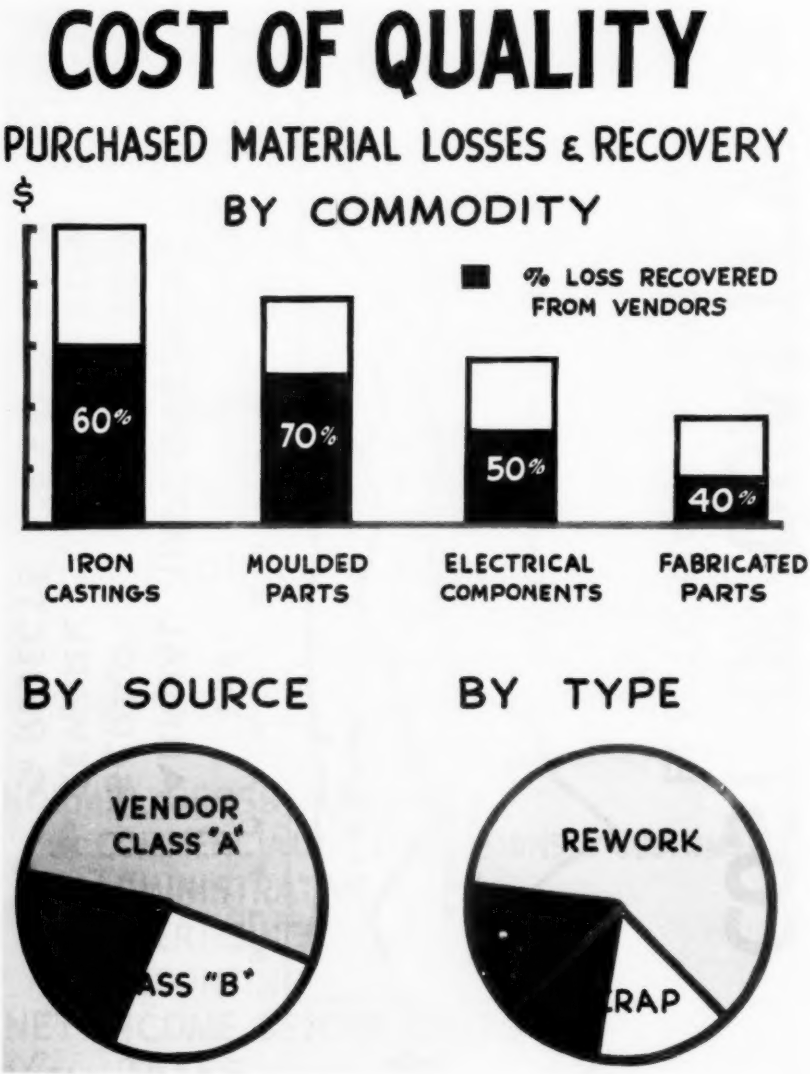

COST OF QUALITY
PURCHASED MATERIAL LOSSES & RECOVERY
BY COMMODITY
$
% LOSS RECOVERED FROM VENDORS
60%
70%
50%
40%
IRON CASTINGS
MOULDED PARTS
ELECTRICAL COMPONENTS
FABRICATED PARTS
BY SOURCE
BY TYPE
VENDOR CLASS "A"
ASS "B"
REWORK
RAP

CHART 5

QUALITY CONTROL IN GARMENT MANUFACTURING

Robert A. Posey
Peter Pan Manufacturing Corp.

A practical working definition of quality control in garment manufacturing is:

> "The application of engineering principles and statistical techniques to all functions of garment manufacturing that affect product quality."

This definition correctly implies that a successful quality control program should actively embrace all levels of management and most departments of the production organization. A sound program must integrate the quality functions of at least five major operating areas of a company. These areas are:

1. Management Organization and Planning
2. Garment Design and Development
3. Raw Material Purchase and Quality Control
4. Manufacturing Operations
5. Analysis of Customer Wants and Complaints

The operating areas act as a link chain. Omission or neglect of any one link will seriously hamper and dilute the potential accomplishments and overall results of the quality control program. Let us further explore the development and application of quality control techniques to these five operating areas.

I. MANAGEMENT ORGANIZATION AND PLANNING

Production and quality are two equal and inseparable partners which cannot survive without each other. To maintain a successful quality program, top management must realize that quality is truly an equal partner and, therefore, deserves full backing and cooperation at all times.

The development of a company quality control program in the garment industry is not an easy task. It must override the skeptics, it must deal with prejudices and must overcome deep-seated traditions and customs. It means the institution of new and seemingly radical ideas, which is never easy to accomplish even under ideal conditions. The person responsible for managing this program must possess three major characteristics in the following order of importance:

A. Ability to get along with all types of persons and the ability to sell and carry out new ideas.

B. Overall knowledge of engineering and quality control methods.

C. Detailed knowledge of product characteristics and garment manufacturing operations.

Once the right person is obtained, he should develop his program and submit it in writing to top management for their approval. Once approved and the program gets under way, management must provide full and unwavering backing at all times.

With regard to the organizational setup, the quality control department is most effective when it acts as a staff function to a top company officer, but maintains close ties with production personnel. This enables the quality control manager to observe, recommend and carry out his program in all company areas that affect quality. It is extremely important, however, that the manager and his staff are not put in a position whereby they can be overruled by line supervision, which may subsequently comprise the immediate position and ultimately the overall program.

II. GARMENT DESIGN AND DEVELOPMENT

All the supervision, inspection and fancy machines in the world cannot correct a poorly designed and engineered garment. It is important to realize that quality is mainly a function of two things: the design and the conformance of units to that design. Basic design and product specifications must be well planned and tested before a formal production run is contemplated. Always allow sufficient time for product development and trial runs so that unforeseen, but costly bugs may be ironed out before committing valuable amounts of men, machines and material. New ideas which seem simple and easily applied may have disastrous effects if not first scrutinized and tried out by a qualified product engineering staff. Top management must have a master planning schedule and a standard procedure for maturing an embryo design into a serviceable, practical product. Necessary steps in this development cycle are:

A. Design of initial garment samples for general top management approval of appearance, cost and merchandising potentialities.

B. Review of proposed garments by a "Product Engineering Committee" composed of members from production, designing, methods engineering, purchasing and quality control departments. This committee should consider the following points:

   1. Scrutiny of all sewing operations and machine attachments with ultimate aim of lower cost and better quality.

   2. Development of most efficient and practical sewing methods and process flow, leading to tentative "operational specifications".

   3. Consideration of type and quality of materials to be utilized, leading to tentative "bill of materials".

C. Factory test runs are necessary to prove out operations, materials, attachments, basic fit and pattern grading. These test runs are a critical component of the development cycle and should never be omitted. The tests may run through a sewing factory under the guidance of responsible supervisors or be processed through an independent product engineering sewing department which specializes in garment development work.

D. Exhausting lab and field tests are needed to prove out serviceability characteristics. In many garments, inherent defects do not show up until used by the customer. Thus, field tests or simulated laboratory tests are important before large scale production runs are shipped. Many textile and apparel companies have experienced an occasional "flop". In many cases, the underlying fault was with management's failure to properly field test their product and to eliminate potential sources of complaints.

E. Final review by "engineering committee" for possible revisions or additions to product specifications is the last step prior to manufacturing.

It cannot be overemphasized that a relatively small amount of development time and effort pays big dividends. Conversely, if development time and techniques are insufficient, heavy losses in money and irrepairable damage to company prestige may result.

III. RAW MATERIAL PURCHASE AND QUALITY CONTROL

Quality control of raw materials actually starts with the quality control system used by the textile and other supplier groups which feed the garment sewing plants. Raw material being processed into garments must be of the right kind, be serviceable, be the correct shade and be of uniform quality from lot to lot. The best way to achieve this goal is to initially know your supplier and make sure that he is as interested in quality control as you are. To accomplish this, we use a variety of scientific tools:

A. MATERIALS DEVELOPMENT

All materials used in your garments should be exhaustibly tested in field and in laboratory. The information acquired through this testing is compared with competitive merchandise; and finally, a product engineering committee consisting of high ranking members of designing, production, sales, purchasing and quality control must approve the materials before any further step is sanctioned.

B. PURCHASING BY SPECIFICATION

Purchase all raw materials by specification. This specification should contain exacting standards required of the material, including every quality characteristic which might conceivably affect the end use and serviceability of the finished garment. When the specification has been completed, it should be reviewed by the management. Prospective suppliers are then invited to a conference in order to familiarize them with specifications. Next, production samples are submitted by suppliers and checked by the laboratory. Finally, the best rated suppliers are selected and copies of specifications and purchase orders are sent to them, together with approved samples. The purchasing department should make every effort to evaluate the prospective supplier from the standpoint of his own quality control abilities. It definitely pays to have the supplier correct his own mistakes rather than run the risk of trying to catch mistakes in your own plant and possibly failing to do so.

C. RECEIVING, INSPECTING AND TESTING

All materials received by a garment manufacturer should be inspected and tested to determine whether or not they adhere to the standards in the purchasing specification. It is at this point that management can profitably apply the techniques of statistical quality control. The following is a typical example of a raw material inspection system:

1. Wide goods are perched by trained examiners who record the weaving defects, spots and stains and other irregularities in accordance with a defect classification chart. At the same time, the length and width of the rolls, identification and quality of packaging are also noted. It is not necessary to examine every roll in the shipment to obtain an accurate quality picture. Instead, only a small number of representative rolls are randomly sampled from the shipment for inspection. The number of defects are compared to an A. Q. L. Chart (A. Q. L. stands for Acceptable Quality Level), which provides an acceptance number for each sample size and quality level desired. If the defects found are equal to or less than the acceptance number, the entire lot is accepted. If the defects found are more than the acceptance number, then the entire shipment is deemed unsatisfactory and is rejected. This inspection technique requires significantly less inspection hours per lot, eliminates 100% screening and forces the supplier to maintain a uniformly high quality product at the risk of receiving back entire shipments.

TABLE I

A	B	C	D
Lot Size Received (Yards)	To Be Examined (Yards)	No. of Pieces to be Sampled	Yards Examined Per Piece
301 to 500	50	1	50
501 to 800	75	2	37 4/8
801 to 1,300	110	3	36 4/6
1,300 to 3,200	150	3	50
3,201 to 8,000	225	5	45
8,001 to 22,000	300	6	50
22,001 to 110,000	450	10	45

Table I shows typical sampling plans used by the Peter Pan Manufacturing Corp. for the sampling and inspection of piece goods. The plans were originally derived from Mil-Std-105-A.

TABLE II

Sample Size Yards Examined	Acceptable Quality Levels							
	1.0	1.5	2.5	4.0	6.5	10.0	15.0	25.0
	Ac. #	Ac. #	Ac. #	Ac. #	Ac. #	Ac. #	Ac. #	Ac. #
50	1	2	3	4	6	9	13	20
75	2	3	4	6	9	13	19	29
110	3	4	6	8	12	18	26	40
150	4	5	8	11	17	24	34	53
225	5	8	11	17	24	34	48	76
300	7	10	14	20	32	44	63	98
450	10	14	20	29	43	62	89	--

Table II is an A. Q. L. Chart which is used by the Peter Pan Mfg. Corp. to accept or reject piece goods shipments.

2. Swatches of material are randomly sampled from each and every incoming shipment and tested in accordance with A.S.T.M. specifications for:

    a. Construction

    b. Weight

    c. Tensile strength

    d. Shrinkage

    e. Elasticity

    f. Washability

    g. Seam efficiency

    h. Permanence of finish

Again, failure of any one quality characteristic to meet prescribed standards may result in rejection of the entire shipment.

3. Effective and surprisingly low cost inspection of a large volume of accessory items is a chore made to order for statistical quality methods. This is accomplished by correctly sampling and inspecting a small number of units in a shipment and accepting or rejecting the entire lot on the basis of the findings in the sampling.

4. Shade control is an important and sensitive operation in the garment industry. Where incoming piece goods are uniform throughout the roll, shades can be inexpensively classified by cutting swatches from each roll of goods and annotating the identification information on the swatches. Swatches should be matched in a special shading room equipped with neutral lights and surroundings. Another even less expensive approach to shade control is to arrange with a reliable supplier to cut swatches from each roll prior to shipment. Thus, the swatches can be mailed to the shading department and the entire shipment shaded before receipt of the materials. Always maintain a shade control record for suppliers. This record should include name of supplier, date of shipment, roll lengths and number of rolls per shade category. A review of this record will prove helpful in planning the size of cutting markers. It will also provide valuable data necessary for weeding out suppliers who cannot supply relatively large amounts of uniformly shaded goods.

5. All inspection data are funnelled to a qualified quality control technician, who has the power to accept or reject the incoming materials. If a shipment is rejected, a "rejection notice" is distributed to the receiving, purchasing and production departments. The purchasing department then takes steps to bill the supplier and return the substandard merchandise.

6. Suppliers should be rated. All testing and inspection data must be compiled in a master book and statistically analyzed each month. As soon as a supplier indicates laxity (i.e. shipping goods inferior to that of his competitors or under the accepted standards), a top management meeting is called to review the quality history record. Here, the ax may fall. In some cases, where warranted, the business is shifted to a supplier who is more reliable and can provide higher quality products.

7. Raw material quality control does not end with the acceptance or rejection of incoming shipments. Materials must be continuously screened during manufacturing operations. Throughout the various sewing and assembling operations, inspectors are on the lookout for defective materials, while the supervisors and staff people evaluate the material from the standpoint of the sewing operators. Imperfect garments caused by defective material must be taken out of the production line. These garments should be closely analyzed and a weekly Imperfect Product Report submitted to management, so that special remedies may be devised and assignable causes eliminated.

IV. MANUFACTURING OPERATIONS

"Gosh, this garment doesn't look or feel like the sample.". These famous last words reflect the thoughts of a buyer or customer when there is an inadequate quality control system. They also mean that good designing and good product development work can be unbelievably distorted without proper controls.

Management will find that by instituting a good quality control system in their manufacturing plants that they will not only assure uniform conformance to initial design, but will also create manufacturing efficiencies, reduce production costs and increase the probability of higher sales and customer acceptance.

Quality of product in the sewing plants can only improve when a dynamic, progressive system is installed which gradually and continually stimulates all manufacturing functions to higher standards of workmanship. Some of the ingredients of a "Dynamic-Progressive" quality control system are as follows:

A. OPERATING PROCEDURES AND QUALITY STANDARDS

The first requisite of good sewing quality control is that all operations performed on a garment and the quality standards expected from each operation must be clearly defined in specification form for use by the various levels of production personnel.

In addition, written standard operating procedures, which can be used as formal directives and training guides, should be put into effect. The use of these tools immediately eliminates guesswork, poor memories, confusion, ambiguities, etc. They are the foundation for a good control system.

B. PATTERN AND CUTTING CONTROL

Accurate cutting is essential. Set up an inspection system in the cutting room to check patterns for wear, accurate grading and notching. Inaccurate and worn patterns and dies are a very large source for quality and production inefficiencies. Cuttings should be identified so that they are easily traceable to the responsible cutter, bundler, lot number, etc. The cuttings should be constantly checked on top, middle and bottom for conformance to pattern. By recording the findings against a cutter's name, a history record can soon be established, which will reveal the persons chronically responsible for bad work. Remember, keep cutting accurate and your sewing problems will be noticeably reduced.

C. OPERATOR TRAINING

The training of sewing operators in fundamental principles of good workmanship and good all-round factory behavior is a basic rule. She must be given adequate time to learn her work and to do a correct quality job before the pressure of piecework production is applied. She must be trained to watch her machine for signs of mal-operation. She must observe good housekeeping rules. She has to learn what to do when bad work is created or discovered. She must follow the correct sewing method. Following is a typical set of rules which must be learned by Peter Pan sewing operators:

I. MACHINES

a. Do not sew on a poorly operating sewing machine. If unsatisfactory, stop machine and immediately call the supervisor. Supervisor will check out machine and, if defective, will call a mechanic.

b. Machines shall be properly oiled and cleaned, daily and weekly, in accordance with plant machine maintenance procedures.

2. HOUSEKEEPING

a. Prior to starting up each morning, dust and clean your work area.

b. After each oiling, run practice cloth through machine to absorb excess oil. Do not leave oil or oil cans on machine tops.

c. At the end of each working day, dust and clean your work area. Cover all material in your area.

d. At the end of each week, all major machine sections shall be properly oiled and cleaned, in accordance with maintenance procedures.

3. QUALITY CONTROL

a. Do not sew over a bad prior operation. If prior operation is bad, stop machine and ask supervisor for a decision.

If rejected, bundle shall be returned to responsible operator for repair.
If accepted, supervisor shall initial the master ticket and work can be resumed.

b. Each operator is responsible for good quality work. During the sewing operation, spot check a few pieces throughout the bundle to make sure that your workmanship and the machine are satisfactory. If not, stop machine and, if necessary, call the supervisor.

c. If you make a bad piece, lay the unit aside until completion of the bundle. Then rip out and repair. Upon receipt of repairs from an inspector, finish the bundle you are working on, then carefully and satisfactorily repair the bad garments. Trim all threads resulting from the repair.

d. If defective material is discovered, lay piece aside and call floor girl for a recut.

4. METHOD CONTROL

a. Each operator shall handle the material and sew with exactly the same method as instructed by the supervisor.

D. SUPERVISOR FORCE

Sewing supervisors must not only be expert sewing instructors, but should know how to perform other functions of good management, such as production control, quality control, personal relations, etc. Too often in the needle trade industry a supervisor is picked for sewing aptitudes alone. They must have the ability to grasp new

ideas, to exert leadership, and get along with persons and operations that come before and after their own. Most of all, if a quality control program is to succeed, the supervisors must be willing to cooperate and have a genuine desire to improve the quality of product.

In addition, a supervisor must really supervise her operators and not get tangled up in a mess of small detail work. If a supervisor does not check her operators constantly and systematically, she cannot effectively control quality.

E. IN-PROCESS INSPECTION

The installation of inspection points after strategic sewing operations is an extremely important ingredient in the quality control system. Both 100 percent inspection and sampling techniques may be applied here with considerable reward. The object of in-process inspection is to find and return to the operator bad work which may later be covered up and which may also affect final quality characteristics. Following is a typical procedure used by the Peter Pan Mfg. Corp. for an in-process inspection point:

1. Immediately after the "Inserting Operation", all bundles of all styles shall be forwarded to "In" storage bins at the first inspection point.

2. The In-Process Examiner shall pick up a bundle from the "In" storage bin and bring it to her work station. She shall then carry out the following general procedure:

   a. While tied, count the number of garments in the bundle. If there are more or less than 60, call the supervisor.

   b. Untie bundle, spot check sizes and dimensions against specifications. If the garments are not correct size, they shall be turned over to a supervisor for further checking and disposition.

   c. Every garment of every bundle shall be thoroughly and carefully examined to determine whether they are perfect or defective. All examiners shall use the same inspection procedure.

   d. Every defect shall be marked with special tape. Garments containing bad workmanship and/or defective material shall be handled in the same manner.

   e. After inspection of each full bundle, defective units shall be separated into groups according to operation.

   f. Each examiner shall maintain an "In-Process Inspection - Repairs Record" and record the following information:

      Inspector's Name
      Date
      Operator's Number
      Type Operation
      Repairable Units per Bundle

   g. Defective units shall be given to a floor girl for dis-

tribution to the responsible operators. The balance of the bundle shall be placed in a "Hold" bin until the defective work is satisfactorily repaired and returned.

Note 1: A small number of repairs per bundle shall be distributed directly to the responsible operator and also picked up by the floor girl.

Note 2: When there is a large number of repairs per bundle (determined by Quality Control), the repairs shall be distributed by the floor girl to the section supervisor, who shall process them and give them to the floor girl for return to the inspection point.

h. Each repaired garment shall be rechecked by the examiner. If the garments are satisfactory, they shall be placed in the bundle. If not satisfactory, they shall be returned to the supervisor. The bundle shall then be placed in the "Out" bin for distribution to the next sewing operation.

There are two important points to remember about in-process inspection. The first point is that the recording of inspection findings is of paramount importance. Proper evaluation of the in-process information will show that, in general, a small percentage of operators create a large percentage of bad work. Thus, by recording and evaluating these data, the plant supervisor has a powerful tool with which to spotlight poor operators and to scientifically watch the results of corrective steps. In time, the garment industry may find that best supervision is not obtained by the familiar round robin patrol inspection, but actually through a process of selective supervision or inspection whereby more supervisory time is deliberately applied to the offenders and less to the good operators. Application of this technique has already shown significant results in the form of less repairs and less irregulars.

The second point to remember about in-process inspection is to make sure that all repairable garments are returned to the responsible operator for her to fix on her own time. By returning work, the operator is automatically made aware of her mistakes and she soon finds that it is more profitable to do a good job the first time.

F. MACHINERY

It goes without saying that proper maintenance of machinery plays a major role in the control of quality. Proper thread tensions, balanced feed dogs, sharpened knives, smooth hooks, correctly adjusted clutches and properly adjusted gauges brought about through "systematic maintenance", assures an army of well trained quality control helpers. It also provides the means for producing quality products at lower production costs.

G. FINAL INSPECTION

The axiom, "Don't let the neighbor (customer) see your dirty wash", certainly applies to the manufacturing of garments and is the basic reason for maintaining a good final inspection department. Through rigid final inspection, defective products are kept from the critical eyes of the customer. It is hard to measure the losses

incurred when defective merchandise is bought by the customer; however, it is safe to assume the losses are usually many times the value of the same product which is rejected at final inspection. Losses are sometimes not immediately apparent but will reflect in long range sales and consumer acceptance. It pays to be critical at final inspection, since the alternative may be lasting damage to the company's sales and hard won prestige. Ingredients of effective final inspection are:

1. Reflection of top management quality policies through the medium of written standard procedures and specifications.

2. Strong supervision and adequate training of inspectors who carry out inspection policies.

3. Adequate recording of imperfect product data for analysis and follow-up by the quality control department.

4. Segregation, recording and return of repairable garments to responsible operators.

5. Independent check of inspectors by quality control department to determine inspection proficiency and conformance to quality control procedures.

Statistical sampling techniques can be profitably used for controlling the inspection proficiency of any final inspection department. One company actually maintains a large trimming, examining department on a piecework basis. Despite the usual weaknesses of piecework inspection, the application of statistical sampling actually creates greater inspection control and at a much lower cost. Following is an actual example of such a system:

1. All trimmed and examined bundles shall be forwarded to a Trimmer-Examiner Checker for final passing.

2. The checker shall separate the taped defective garments from the "perfect" ones. She shall then draw at random a specified number of "perfect" garments.

3. The sample size or number of units to be checked shall be determined by the quality control department and will be based upon the prevailing production and quality performances of the plant.

4. Each unit in the sample shall be carefully inspected for quality of trimming and for any unmarked material or workmanship defects. Any unmarked defects or bad trimming shall be taped and the garment set aside.

5. Upon completing the inspection of all units in the sample, count the number of defective units found. If the number of bad units is equal to or below the specified acceptance number, then pass the full bundle. If the amount of bad garments found in the sample is above the acceptance number, then reject the bundle.

6. Accepted bundles shall be forwarded to the Repair-Segregating Section.

7. Rejected bundles shall be returned to the responsible trimmer-examiner who must immediately reinspect her bundle and tape or

trim the defective garments she missed.

8. The retrimmed and re-examined bundle shall again be forwarded to the checker who processes the bundle in the same manner as before.

9. Each checker shall maintain a "Trimmer-Examiner Quality Record" and record Trimmer-Examiner stamp number, bundle serial number, style number, type defects found and number of defective units found. This information shall be utilized for determining the quality performance of each Trimmer-Examiner and for determining the production of each checker.

The above system provides three vital forces which are dynamic, as well as practical. The first force requires the return of poorly inspected garments to the piecework inspector who, in turn, is monetarily penalized for not doing her job correctly. The second force requires that a record be kept of the checkers' findings. In this way, unreliable inspectors are soon found and either corrected or weeded out of the department. The third force is the sampling plan and its affect on the entire inspection force. Following is a typical plan:

	Units in Bundle	Units Sampled	Acceptance Number	Rejection Number
Sewing Workmanship	60	15	1 Repair	2 Repairs
Trimming	60	15	1 Thread	2 Threads

It is readily seen that by manipulating the acceptance and rejection numbers, the management can effect the average quality level of product going to the customer. Thus, by using these sampling plans, management has a powerful, flexible, yet practical instrument for controlling final quality.

H. IMPERFECT PRODUCT REPORT

One of the most potent tools for solving quality control problems is the Imperfect Product Report. This report serves as a sensitive measurement mechanism for depicting relative quality levels of product types and individual defects found in the finished garments. Briefly, the Imperfect Product Report is set up as follows:

1. Irregular garments are classified as to the exact assignable cause for being defective. *

* The following card is used by the Peter Pan Finishing Department to record the disposition and assignable causes for all defective units in a bundle of 60 garments. At the end of each week, the cards are forwarded to a Quality Control Statistician who uses the data for computing the weekly Imperfect Product Report.

BUNDLE QUALITY ANALYSIS						QUALITY CONTROL	REPAIRS AT PLANT				
STYLE 38-3		SIZE & CUP 34 B		COLOR W		SEWING ORDER NO. 1738			BUNDLE NO. 1223543		
PLANT Utica		DATE TO SINGLES 4/ 18 /54				NO. REPAIRS TO PLANT —			DATE TO PLANT / /		
NO. OF DOZEN REGULARS 4 6/12		QUANTITY DELIVERED TO SINGLES DEPT. IRREPAIRABLE SINGLES 4				REGULAR SINGLES 2			NO. OF SINGLES BEING HELD		
NUMBER OF IRREPAIRABLE GARMENTS AND CODE NUMBERS											
AMT.	CODE	AMT.	CODE	AMT.	CODE	AMT.	CODE	AMT.	CODE	AMT.	CODE
2	5 R	1	18 R	3	12 R						
DEFECTIVE OPERATIONS											
FINISHING DEPARTMENT				SEWING DEPARTMENT							
AMT.	OPER. #	AMT.	OPER. #	AMT.	OPER. #	AMT.	OPER. #	AMT.	OPER. #	AMT.	OPER. #
				2	74	1	201	1	42	2	93
WRITTEN BY WR				RECEIVED IN SINGLES DEPT. 4/18/54				REMARKS			

2. At the end of a predetermined period, the percentage of total finished production versus number of irregulars is determined. In addition, the number of irregulars in each defect classification is compared with total production, which results in percent defective for each assignable cause.

3. In addition to a period compilation, all data are cumulated, thus providing long range quality picture, as well as a periodic one.

4. The percentage data are then compiled into an Imperfect Product Report and distributed to responsible production and technical departments for follow up and correction.

Through the use of this accurate performance barometer, quality trends and significant sources of bad work can be spotted, analyzed and corrected. The Imperfect Product Report will not only show the initial level of bad work, but will accurately indicate the results of successful quality control corrective action, which, if wanted, can be computed into dollars and cents saved.

V. ANALYSIS OF CUSTOMER WANTS AND COMPLAINTS

This garment is defective----I payed for first class merchandise----I want my money back----Wait till my friends hear about this----I'll never buy from that company again----. Thus, another irate and dissatisfied customer chips away at the quality reputation of a company. A potential friend and future sales have been lost. Can this not-too-uncommon scene be prevented or controlled? The answer is YES! But, first of all, the reactions and wants of these customers must be accurately determined. What are the reasons for dissatisfaction?

The basic mechanism for ascertaining customer reactions is the "Imperfect Product Claims Report". This report is developed as follows:

1. All returning merchandise shall be segregated into two major groups:

A. Returned garments containing defects created by manufacturing operations.

B. Returned goods - other reasons.

2. The returned defective garments shall be inspected and classified as to the exact cause and nature of the defect.

3. The number of defective units in each classification accrued over a certain period (usually a month) shall be compared with average sales shipments for the same period. Thus, percent defective for each type defect may be derived.

4. In addition, the claims data may be cumulated and compared with cumulated sales shipments to obtain a wider picture of customer complaints.

5. Upon completion of this report, the quality control department and top management have in their possession valuable clues and an accurate basis for developing a program leading to product improvement and lower customer returns.

With regard to customer complaints, the claims policy and investigation procedures of the company should be spelled out formally. Issue a booklet to field representatives and buyers, stating claims policy. A booklet should be distributed covering the following points:

1. Introduction and Purpose

2. Procedure for Investigation and Reporting of Complaint

3. General Settlement Procedure

4. Claims Description and Disposition Policies.

Take a definite stand on each and every type of defect found. Is or isn't the company responsible? In this way, company representatives can take a stand and resolve differences with a minimum of friction and communication. With specific descriptions of each defect, the right picture can be transmitted to the plant for effective analysis and follow-up.

Major weaknesses found in the claims system of some companies stem from lack of accurate information or standard field nomenclature. Another weakness is the excessive time lapse between complaint and satisfactory conclusion of the complaint. The use of formal policy eliminates these basic claims weaknesses.

As an additional instrument for detecting the ground swell of customer opinion at "grass roots" level, representatives and stylists in the field should periodically visit stores in all parts of the country. In the stores, they spot check the merchandise and interview customers. A detailed report on customer reaction, including complaints as well as favorable testimonial, is then dispatched to the quality control department for review and follow-up. The analysis of customer complaints and the interview of customers in the field are a form of consumer research which, in turn, is an integral part of the manufacturing operations. Consumer research acts as a servo-mechanism which, by probing into the reasons for customer preferences and dislikes, yields hypothesis that

assist management in making intelligent decisions regarding changes in design, quality uniformity and production levels.

Consumer research is basically the development of organized communication between the manufacturer and the users, or potential users, of his products.

Thus, if the customer has a chance to get across his point of view and if this viewpoint is resourcefully integrated into the manufacturing operations through a well rounded quality control system, then the probability is high that the customer will receive a better product, better suited to his needs and at a lower cost.

Process Capability Considerations in Product and Process Design

Leo J. Jacobson
International Resistance Company
Philadelphia 8, Penna.

This presentation deals with the contributions of the methodology of modern quality control in the design of products and processes. The advantages of early use of process capability studies in the process design and planning stage are presented. Particular emphasis is given to the analysis of cost implications of excessive reduction of process variability. The impact of recent developments in automatic gaging and sorting equipment in providing the basis for deliberate design of high variability processes is examined. The value of variability studies in selection and design of metering and test equipments is discussed. Some applications - correct and incorrect - of "feed-back" process control systems are analyzed. Illustrations taken from the speaker's experience in electronic component manufacture have been selected with a view toward requiring a minimum, specialized technical or engineering knowledge.

STATISTICAL DESIGN OF EXPERIMENTS IN METALLURGICAL RESEARCH

S. Gilbert
Research and Development Laboratory,
United States Steel Corporation

It has been stated by R. A. Fisher[1)*], "Critics who refuse to accept the conclusions supposedly proved by experimental evidence are accustomed to take one of two lines of attack. Claims may be made that the interpretation of the experimental data is faulty. Such criticisms of interpretation are usually treated as falling within the domain of statistics. However, aside from the responsibility of the statistician to understand the processes he applies or recommends, the basic understanding of scientific inference is the obligation of all concerned with the experiment."

A second type of criticism to which experimental results are exposed is that the experiment itself was poorly designed. Many research programs from which decisive conclusions have been drawn were based on high-powered interpretations of very meager data. The essential point made by Fisher[1)] is that both types of criticism are aimed at different aspects of the same thing. If the design of an experiment is faulty, any method of interpretation which makes it out to be decisive must be faulty, too. Statistical procedure and experimental design are two different aspects of the same requirements for successful experimentation.

The processes by which steel is manufactured and fabricated lend themselves readily to the application of certain statistical techniques in analyzing the problems presented to the Research and Development Laboratory. Whether in estimating the average particle-size distribution of a particular grade of iron ore from the Mesabi Range, or in determining the long-term corrosion of piling in sea water, certain observations of these phenomena must be made before they can be evaluated. The line of attack that a research worker generally follows in evaluating a given process is to formulate hypotheses about the process, and then to proceed to disprove or verify these hypotheses in the light of the observational data that he has collected. The purpose of experimental design is to ensure that the experimenter obtains adequate data relevant to his hypotheses in as economical and efficient a way as possible. It is in this phase of experimentation that the statistician can make a major contribution to the research worker.

Experimentation may be divided into two broad subclasses: absolute and comparative. An absolute experiment is one in which the exact value of some measurable quantity is to be determined. An example of this type of experiment in the steel industry would be the determination of the average particle-size distribution of a particular grade of iron ore shipped to a blast furnace. A comparative experiment is one in which two or more treatments are compared in their effects on a chosen characteristic of a population. If an experiment is conducted to determine the effect of two different types of ore on the iron production of a blast furnace, we say that a comparative experiment is being run.

Experimentation in the laboratory, where the observations are made with control effected by the experimenter fixing the levels of the variables, or by statistical control effected by randomization, is

generally of the comparative type. However, in conducting a sample survey in the field, as is often done, the concepts of absolute experimentation are abandoned when the experimenter finds that to fix the value of some quantity, such as the average size-consist of a grade of iron ore, he must first investigate the importance of the factors that influence the size of the ore as it comes to the blast furnace. This leads him into a comparative study, even though he cannot control the variables affecting the quantity he is observing. An example of an uncontrolled comparative study occurred when the Research and Development Laboratory undertook to estimate the degradation of iron ore being shipped from the mines in Minnesota to the blast furnaces in the Pittsburgh area.

The problem was to determine the degradation, or breaking up, of the ore due to handling in shipment. Samples of the ore were obtained from the screening plant located at the mines in Minnesota, and from the bins adjacent to the blast furnaces in Pittsburgh. In between the two sampling locations, the handling of the ore could not be controlled by the men conducting the test. It was handled in a standard manner representative of all ore shipments. A number of samples, from which an estimate of the average size-consist could be derived, were obtained at each sampling area.

The measurement data obtained in this sampling program enabled the experimenters to evaluate the magnitude of the ore degradation between the point of shipment and the point of actual use. The estimate of the average difference in size-consist between these two points, and the statistical uncertainty associated with estimating this average difference, were then used in determining the optimum location of a particular ore-treating plant.

More frequently, however, the experiments conducted at the Research and Development Laboratory are of the controlled comparative type. In these test programs the application of statistical design to the problem has led to more efficient collection and analytical use of test data, and to greater validity of the conclusions drawn from them.

Frequently the experimenter is faced with the problem of determining the effect that the variation of several factors will produce on a particular process. To facilitate efficient experimentation, two concepts of experimental design are introduced into the plan: multiple balance of experimental conditions, and randomization of experimental trials. First, the concept of multiple balance of experimental conditions, introduced by R. A. Fisher[1)], is employed. Its application is illustrated in an experiment involving the evaluation of the effects of two variables, call them A and B, on a particular physical property of steel. The classical and now outdated concept of varying only one factor at a time and comparing its result with that of the control level, would lead to an inefficient estimation of the effects of the variables on the quantity being measured.

The factors, A and B, are to be studied at two levels. The lower level of each factor can be denoted by 0 and the upper level by 1. To investigate these factors at the two levels in the classical manner would require making only three runs. The three runs and the level of each factor in the runs are shown in Table I.

Table I

Factor	A	B
Run 1	0	0
Run 2	1	0
Run 3	0	1

The three runs in Table I will suffice to determine the effect produced either by changing factor A from its lower level to its higher level, or by varying factor B. This is accomplished by comparing Run 1 with 2 and Run 1 with 3, respectively. However, these three runs are unbalanced in two respects. Of the six factor levels, four are at the lower level and two are at the higher level. For each comparison the control run, Run 1, is used to make a judgment of the effect produced by varying the levels of A and B. Furthermore, these runs furnish no information on the effect produced by factors A and B both being present at the upper level. It is also conceivable that the effect produced by changing the levels of factor A from low to high would not necessarily be the same at both levels of factor B. This is known as "interaction", which can be defined as the failure of one variable to produce the same magnitude of effect at all levels of another variable. At times the evaluation of the interaction term of an experiment is of more interest than the determination of main effects. Consequently, when the existence of an interaction is likely in an experiment, the experimental design must be constructed to enable determination of this interaction.

The set of runs in Table II will enable the experimenter to evaluate the main effects of factors A and B, and the A x B interaction for this program.

Table II

Factor	A	B
Run 1	0	0
Run 2	0	1
Run 3	1	0
Run 4	1	1

A comparison of this set of four runs with the set outlined in Table I shows several advantages. First, the statements made concerning the effects of varying factors A and B are more precise. In the design proposed in Table II the main effects are obtained by comparing the two runs made at the lower level of the factor with the two runs made at the higher level. Thus all four runs are used to measure each main effect, by this means as much precision being attained in each conclusion as if all four runs were devoted solely to the measurement of a single factor. The A x B interaction is determined from a comparison of the difference in results of Runs 1 and 4, with the difference in results of Runs 2 and 3. All four runs have also been used in determining this relationship. Thus from four runs three conclusions have been derived, each conclusion

being based on experimental evidence supplied by all four runs.

The two-factor experiment described above is the simplest case of an experimental plan known as the factorial design. The interaction is the important effect about which the factorial design can give information. However, even in the analysis of a problem where it is known that an interaction is nonexistent, the factorial design does not lose any information. If it can be assumed beforehand that the interaction of factors A and B is nonexistent in the design laid out in Table II, this design still furnishes more information about the main effects produced by the two factors. The average difference between the two levels of factor A is obtained from two individual differences, and likewise for the average difference between the two levels of factor B. Thus, in the absence of an interaction, the factorial design furnishes a certain amount of "hidden replication", or duplication, depending upon the number of factors being investigated.

After completion of the program to determine the effects produced by factors A and B, it was thought that a third factor, C, might have some effect on the property being determined. It could not be assumed that the effect of factor C would be entirely additive, that is, would not interact with the two other factors. An expansion of three factors each studied at two levels required that eight runs be made for a fully balanced array. These eight runs would furnish sufficient information to arrive at seven conclusions on the effects produced by the variation of three factors. Of these seven conclusions three would be on the main effects of the factors A, B, and C; three on the failure of each of the single factors to produce the same effect at both levels of another factor (first order interactions); and one on the failure of the A x B interaction to be different at the two levels of factor C. This would be the A x B x C interaction. The same quantity also measures the difference of the A x C interactions at the two levels of B, and similarly for the B x C interaction. An interaction of three factors such as this is termed one of the second order.

The factor levels of the eight runs to be made are given in Table III.

Table III

Factors	A	B	C
Run 1	0	0	0
Run 2	0	1	0
Run 3	1	0	0
Run 4	1	1	0
Run 5	0	0	1
Run 6	0	1	1
Run 7	1	0	1
Run 8	1	1	1

Close examination of the factor levels of the eight runs reveals that the makeup of the first four runs in this set, with the exception of the added C factor, is identical with the four runs listed in Table II. Consequently two possibilities were open to the experimenter at this point. If the first four runs, which were made to evaluate the effects produced by factors A and B, were free of any systematic bias (for example, the effect of seasonal variation affecting the quality of raw materials entering the process, or any other nonrandom fluctuations occurring over the time interval between the completion of the first experiment and the running of the second), then in the investigation of the effect of the third factor, the data from the first four runs with C at its lower level, as in Table II, could be used in conjunction with Runs 5 through 8. If, on the other hand, a systematic bias due to the time interval between experiments was suspected, then, to arrive at an unbiased estimate of the effects produced by the three factors, the full set of eight runs in Table III would have to be made. In addition, if a trend due to the order in which these eight runs are made is likely, then a random sequence of making the runs must be chosen to minimize the associated time effect. This second important aspect of experimental design—the concept of randomization of experimental trials—is too often passed over lightly by the experimenter. An attempt should be made to randomize all experimental conditions that are not under the control of the experimenter but that might produce an effect on the quantity being measured. The simple precaution of randomization guarantees the validity of the test of significance by which the result of the experiment is to be judged. It is not in the scope of this paper to offer more than a brief reference to this important concept.

The number of runs required by any factorial design can be determined by the simple expansion of a term, p^n, where n represents the number of factors being varied and p denotes the number of levels at which each of these factors are to be studied. For simplicity, the cases dealt with thus far in this paper have been restricted to two levels. However, the general factorial design is not restricted to the investigation of two-level factors only. The theory, construction, and practical examples of factorial designs of a more complex nature are well covered in all the references appended to this paper.

The relative complexity of some of the processes being evaluated at the Research and Development Laboratory is such that the fulfillment of the concept of balance, given by the factorial design, leads to the making of far too many experimental runs. In an experiment to evaluate the effects of eight factors, each being varied over two levels, a total of 256 runs would have to be made. Even with the very simplest experimental device for making these runs, the experimenter will often balk at making such a prodigious number of tests. From an economic and practical viewpoint, it is sometimes virtually impossible to test the number of treatment combinations imposed by the use of the complete factorial design of experiment. Furthermore, when the number of factors is greater than five or six, the great majority of the information furnished by the factorial design relates to high-order interactions, which are unlikely to be either real or of interest.

In an experiment involving the evaluation of the effects produced by eight factors at two levels each, it was reasonable to assume that the experimental error was small and that the high-order interactions involving 4 or more factors were negligible. The pooling of these high-

order interactions, which account for 163 out of the possible 255 degrees of freedom for this experiment, furnished a well estimated term for use in judging the significance of the effects that were to be evaluated. When experimental error is small and the high-order interactions are not of appreciable magnitude, the precision furnished by the complete factorial design is unnecessary. It is possible to carry out only a fraction of these large factorial experiments and still retain satisfactory information on the main effects and lower-order interactions. The problem confronting the experimenter is which of the total number of runs for the full factorial shall be included in the partial or fractional factorial set to be run.

The answer to experimental situations of this kind involves the use of "Fractional Replicate Designs". This type of design was first reported by Finney[2,3]. The construction of these designs is well covered in statistical texts by Brownlee[4], Kempthorne[5], and Cochran and Cox[6]. A paper by K. A. Brownlee, B. K. Kelly, and P. K. Loraine[7] presents a method for constructing fractional designs having as many as 512 runs. For the industrial research worker who cares little about the mathematical background of these designs, the article by O. L. Davies and W. A. Hay[8] is a good introductory work in the construction and uses of fractional factorial designs. It is not within the scope of this paper to present the underlying assumptions and theory of the fractional replicate design. Instead, the utility of the method will be illustrated by an example of its application to an industrial research problem.

A program was initiated at the Research and Development Laboratory to study the effects of some of the operating variables that influence the performance of an open-hearth furnace. The test program was to be carried out on a pilot-plant level in a model furnace of the same design as a full-scale operating open hearth. Since the primary purpose of the study was to determine the "firing" characteristics of this particular furnace design, it was not necessary to actually melt and refine steel in the model. Instead, a complete thermal instrumentation of the model furnace was made. The instrumentation permitted the measuring and recording of such things as fuel rate, air rate, Btu to the hearth, and Btu lost in the waste gases, all of which affect the efficiency of an open-hearth furnace. The furnace was also instrumented for the precise control of the experimental conditions of operation whose effects were to be evaluated. Although the physical equipment necessary for running the test was available, the proper experimentation for its most efficient use had to be designed.

Eight operating variables, which were thought to be of primary importance to the efficient firing of an open hearth, were chosen to be studied at each of two operating levels. Each variable is called a factor and the various states or conditions of that factor are called its levels. For example, furnace pressure would be a factor and the actual pressures used would be the levels of that factor. The factors chosen must be capable of being varied over the levels independently of the levels of the other factors in a particular experimental run. Since the aim of this example is to illustrate a method rather than to report specific information on the firing of an open hearth, it will suffice to refer to the eight factors being studied as A, B, C, D, E, F, G, and H. Similarly the actual operating levels of these factors can be denoted by a zero (0) for the lower level, and a one (1) for the upper level of a particular factor.

The full investigation of the effects produced by the variation of

eight factors from one level to another, in the context of all possible combinations of the other seven, would require making 2^8 or 256 runs. To make this many runs, involving both time and materials, would be uneconomical, impractical, and, as will be shown, unnecessary for the scope of this program.

From previous information on the behavior of the model furnace, it was concluded that the experimental error associated with the measurements made was not appreciable. Furthermore, it could reasonably be assumed that the magnitude of the high-order interactions would be small, and that only a vague meaning could be attached to them. On the basis of these assumptions, a balanced selection of a fraction of the total number of runs required by a full factorial design could be made. It was arbitrarily agreed that 64 runs could efficiently be made; this number of runs would furnish information on each of the eight main effects and each of the 28 first-order interactions. In making only 64 runs out of a total of 256 required by the full factorial, the amount of experimentation was reduced by 75 per cent, the price being paid for this reduction being the inability to estimate the effects of the higher-order interactions. The 64 experimental runs with the levels of each of the eight factors for the runs are outlined in Table IV.

Table IV

	FACTORS									FACTORS							
Run	A	B	C	D	E	F	G	H	Run	A	B	C	D	E	F	G	H
1.	1	1	1	1	0	1	1	0	33.	0	1	1	1	1	1	1	1
2.	1	1	1	1	0	1	0	1	34.	0	1	1	1	1	1	0	0
3.	1	1	1	1	0	0	1	1	35.	0	1	1	1	1	0	1	0
4.	1	1	1	1	0	0	0	0	36.	0	1	1	1	1	0	0	1
5.	1	1	1	0	1	1	1	0	37.	0	1	1	0	0	1	1	1
6.	1	1	1	0	1	1	0	1	38.	0	1	1	0	0	1	0	0
7.	1	1	1	0	1	0	1	1	39.	0	1	1	0	0	0	1	0
8.	1	1	1	0	1	0	0	0	40.	0	1	1	0	0	0	0	1
9.	1	1	0	1	1	1	1	0	41.	0	1	0	1	0	1	1	1
10.	1	1	0	1	1	1	0	1	42.	0	1	0	1	0	1	0	0
11.	1	1	0	1	1	0	1	1	43.	0	1	0	1	0	0	1	0
12.	1	1	0	1	1	0	0	0	44.	0	1	0	1	0	0	0	1
13.	1	1	0	0	0	1	1	0	45.	0	1	0	0	1	1	1	1
14.	1	1	0	0	0	1	0	1	46.	0	1	0	0	1	1	0	0
15.	1	1	0	0	0	0	1	1	47.	0	1	0	0	1	0	1	0
16.	1	1	0	0	0	0	0	0	48.	0	1	0	0	1	0	0	1
17.	1	0	1	1	1	1	1	1	49.	0	0	1	1	0	1	1	0
18.	1	0	1	1	1	1	0	0	50.	0	0	1	1	0	1	0	1
19.	1	0	1	1	1	0	1	0	51.	0	0	1	1	0	0	1	1
20.	1	0	1	1	1	0	0	1	52.	0	0	1	1	0	0	0	0
21.	1	0	1	0	0	1	1	1	53.	0	0	1	0	1	1	1	0
22.	1	0	1	0	0	1	0	0	54.	0	0	1	0	1	1	0	1
23.	1	0	1	0	0	0	1	0	55.	0	0	1	0	1	0	1	1
24.	1	0	1	0	0	0	0	1	56.	0	0	1	0	1	0	0	0
25.	1	0	0	1	0	1	1	1	57.	0	0	0	1	1	1	1	0
26.	1	0	0	1	0	1	0	0	58.	0	0	0	1	1	1	0	1
27.	1	0	0	1	0	0	1	0	59.	0	0	0	1	1	0	1	1
28.	1	0	0	1	0	0	0	1	60.	0	0	0	1	1	0	0	0
29.	1	0	0	0	1	1	1	1	61.	0	0	0	0	0	1	1	0
30.	1	0	0	0	1	1	0	0	62.	0	0	0	0	0	1	0	1
31.	1	0	0	0	1	0	1	0	63.	0	0	0	0	0	0	1	1
32.	1	0	0	0	1	0	0	1	64.	0	0	0	0	0	0	0	0

The number of runs to be made constituted a 1/4 replicate of a 2^8 factorial design. The job of choosing the particular 64 runs out of the possible 256 has been simplified by the work of Brownlee, Kelly, and

Loraine[7]. The information summarized in Table 3 of their paper was used to lay out the fractional replicate for this investigation.

Inspection of the makeup of the runs in Table IV reveals the intricate balance of treatments resulting from the particular quarter fraction chosen to be made. There are 32 zeroes and 32 ones in each column; that is, each factor was investigated at its lower and upper levels in an equal number of runs. For any set of 32 zeroes, (or ones), there are 16 zeroes and 16 ones in each of the other columns, but always a different split for each pair of columns. The requirement of maintaining a balance of experimental conditions for efficient use of all the data collected in an investigation is accomplished with the fractional replicate design. The balance of experimental conditions furnishes information on the effect of each factor in a variety of contexts, permitting a more general use of the results. Furthermore, the balanced set of experiments, particularly when dealing with two-level factors, affords an easy analysis of the resulting data.

It has been shown previously in this paper that for any two factors, call them P and Q, three conclusions may be obtained from four balanced runs. If the experimental conditions for the four runs are P_0Q_0, P_0Q_1, P_1Q_0, and P_1Q_1, the three conclusions may be obtained by putting the four values for the runs in a "2 x 2" table, as follows:

		P 0	P 1
Q	0	P_0Q_0	P_1Q_0
	1	P_0Q_1	P_1Q_1

The effect of changing the level of factor P from its low to its high level is given by the difference in the column averages. The effect of changing from Q_0 to Q_1 is given by the difference in the row averages. The P x Q interaction, which measures the failure of the effect of P to be the same at the two levels of Q, is given by the difference between the two diagonal averages. These three effects are independent, even though each conclusion uses all four results. The expansion from two factors to eight factors does not lead to a more complex analytical solution of the data.

The 36 conclusions (8 main effects and 28 first-order interactions) for the open-hearth program are obtained by merely putting the data into the appropriate four cells in 28 such "2 x 2" tables. Each of the four cells in any two-way table will contain the average of the 16 runs made which fit the cell specification. For example, two of the factors studied were Fuel Temperature and Steam Pressure. Of the 64 runs made, 32 were made with a low fuel temperature and 32 with a higher fuel temperature. Of the 32 runs at a low fuel temperature, half were made at a low steam pressure and half at a higher steam pressure. The same pattern of symmetry existed for the 32 runs made at the high fuel temperature. To estimate the effects of varying the fuel temperature and fuel pressure on the Btu's transmitted to the hearth, the data were entered in the following table*:

---

* The data in this table are not indicative of actual operating data and no significance should be attributed to the trends indicated.

		Fuel Temperature		
		Low	High	Average
Steam Pressure	Low	40	80	60
	High	24	60	42
	Average	32	70	

Each number in the body of the table represents the average of all the data for 16 runs. The row and column averages represent the data from 32 runs each. The overall difference between column averages indicates an increase in Btu's to the hearth when the fuel temperature is increased from its low to its higher level. Similarly, there is an indication of a decrease in Btu's to the hearth when the steam pressure is increased in the range specified by the low and high levels of this factor. The increase in Btu's to the hearth with increasing fuel temperature is somewhat greater at low steam pressure than at high steam pressure. The measure of this failure of the fuel-temperature change to produce the same effect at low and high steam pressure is obtained by comparing the difference between the two increases just mentioned. A similar analysis is made for each of the 27 remaining two-way tables. All the measured data from the 64 runs is used in drawing each conclusion on the main effects and interactions for this investigation. Furthermore, only 36 (8 main effects and 28 first-order interactions of one degree of freedom each) of the total of 63 degrees of freedom for the experiment have been used in obtaining the desired information. This leaves 27 degrees of freedom for estimating the "residual error" term to be used in making significance judgments. The information on the residual error can be used in calculating "confidence intervals" for estimating the average difference produced by changing the levels of a factor[4]. In making these 64 runs, more information has been furnished on the behavior of the eight factors than could have been obtained by a far greater number of actual runs taken just as they occurred in practice.

The particular fractional replicate design described is only one of several that could have been used to evaluate the effects produced by the variation of the eight factors. This application was chosen because of the simplicity of the design. It is not implied that all fractional replicate designs can be applied as easily as this. The use of a fractional design permits the experimenter to investigate the behavior of a number of factors, even though the number of experiments that he can run is limited. This concept of experimental design can be a very valuable tool in evaluating the complex processes that occur in industrial research. The choice of a fractional design does, however, require considerable care, and the experimenter would be unwise to use this type of design at any available opportunity without first examining all the facets. The fractional replicate design furnishes an experimental program for evaluation of main effects and low-order interactions with a precision of results impossible to attain by any other method. If the factors chosen to be studied are the decisive ones, this design enables their evaluation to be made with a minimum of time involved in experimentation. However, the fractionally designed experiment is very sensitive to missing data and therefore, the interpretation of results cannot be achieved until all the runs have been completed and all the data have been collected. A study of the references given previously will provide a more detailed coverage of the concepts of fractional replication

with practical examples of its use in industry and research.

In both types of experimental design discussed in this paper, the concepts of balance and randomization of experimental conditions have been shown to be fundamental in enabling the experimenter to draw valid conclusions from his experimental data. Neither of these statistical principles for experiments can be applied without the appropriate forethought and planning. They can almost never be superimposed on a mass of data resulting from a number of experiments not previously designed to answer specific questions.

Too often the experimenter seeks the advice of a statistician in making inferences from the results of experiments. Since the inferences that can be made depend on the way in which the experiment was carried out, the statistician can only rarely guarantee a successful interpretation of data gathered without statistical plan. It sometimes occurs that no inferences can be made, or that those which can be made do not answer the questions to which the experimenter had hoped to find answers. The statistician can only indicate how to avoid these outcomes in future experiments. As a result of such unhappy circumstances, it has come to be realized that the time to think about statistical inference, or to seek advice of the statistician, is during the planning of the experiment.

Generally, too little time and effort is devoted to the planning of an experiment. Experimentation is both costly and time consuming. Therefore, it behooves the research worker to consider an intensive planning period prior to the actual performance of the experimental work. It is a good practice to make a written draft of the proposals for any experiment. In a recent acticle by C. A. Bicking[9)], a detailed "check list" for planning test programs was presented. Adherence to a check list, of which the experimental design is a vital part, guarantees the performance of an efficient experimental program to furnish data relevant to the objectives under consideration.

References

1. R. A. Fisher, The Design of Experiments, Hafner Publishing Company, Inc., New York, 1949.

2. D. J. Finney, "The Fractional Replication of Factorial Experiments", Annals of Eugenics, 12, pp 291-301, 1945.

3. D. J. Finney, "Recent Developments in the Design of Field Experiments: Part III Fractional Replication", Journal of Agricultural Science, 36, pp 184-191, 1946.

4. K. A. Brownlee, Industrial Experimentation, Chemical Publishing Co., New York, 1950.

5. O. Kempthorne, Design and Analysis of Experiments, John Wiley and Sons, New York, 1952.

6. W. A. Cochran and G. M. Cox, Experimental Designs, John Wiley and Sons, New York, 1950.

7. K. A. Brownlee, B. K. Kelly, and P. K. Loraine, "Fractional Replication Arrangements for Factorial Experiments With Factors at Two Levels", Biometrika, vol XXV, Parts III and IV, December 1948.

8. A. L. Davies and W. A. Hay, "The Construction and Uses of Fractional Factorial Designs in Industrial Research", Biometrics, vol 6, No. 3, September 1950, pp 233-249.

9. C. A. Bicking, "Some Uses of Statistics in the Planning of Experiments", Industrial Quality Control, vol X, No. 4, January 1954, pp 20-24.

BUILDING QUALITY IN THE PRESS ROOM

William A. MacCrehan
Bendix Radio Division of
Bendix Aviation Corporation

Every pressman is conscious of quality printing . . . Why! Because the basic apprenticeship training of a pressman spends many, many hours of teaching fit of plates, depth of cylinder packing, sharpness of register, etc. Controlling quality then is not a task of teaching the press operator how to do his job, but to make him conscious of trends, cycles and changes in his process. This apprenticeship, which requires a minimum of three years of training to learn a highly skilled trade, must be clearly recognized in dealing with the pressman. It enters into the picture when hourly rate inspectors (usually women) evaluate a printed sheet, when color is under discussion involving the mystic term of "dryback," and when questions arise on uniformity of printing surfaces. A fundamental fact then is that the process printing field still places a high premium on the skill of the worker to control the finished results . . . this is an "art" that is slowly becoming a science.

Where quality control applies to a process involving arbitrary opinions (varying with the each individual's background, training, and experience) the first step is to obtain <u>facts</u> and the second step is to believe the <u>data</u>, <u>not</u> <u>the</u> <u>people</u>.

To gather facts is a three-step program:

1. To clarify and define the press problems <u>controllable by the pressman</u> as opposed to mill stock creating trouble beyond the pressman's control

2. To establish visual standards for use by the inspectors who are unskilled in press-room lore . . . this gives a common ground for mutual evaluation of quality by the pressman and the inspector

3. To reach a sound basis for action with management when the <u>facts</u> indicate lack of control

The first of these factors, definition of responsibilities, can be organized with (1) a frequency of fault chart, (2) a meeting with mill suppliers and unilateral action by a control committee of the plant manager, the purchasing agent, the press room superintendent, and quality engineer.

The fault chart will specifically spell out the symptoms of the defects . . . the lost press time and the action taken to remedy the fault.

The control committee must decide on reviewing the case problem where responsibility for correction exists. This can run the gamut from blending a stock ink to alleviate "tack pull" on surface fibers . . . to outright rejection of the mill stock for "hickies" and "fish eyes." It is at this point that quality control often encounters passive resistance to promoting press room confidence that correction of troubles can and will be the result of cooperative efforts. In some instances the

same problems have been battled over and over again with a resultant effect that an apathetic feeling exists that "Nothing can be done." A further problem is the press room general feeling that the mill product represents the bulk of the problems. (A similar feeling exists on the part of the mill that the press room useage of mill stock represents the bulk of the problems.) . . . The review of case problems from the frequency of fault data will readily separate fact from fiction.

The second phase of the quality attack to press room problems is . . . reaching a common understanding on what constitutes a defect. Here, fertile ground exists for it must be recognized that the press room is basically a job shop, subject to the relentless demands of deliveries keyed to sales programs of the customer. Further, it is imperative to recognize that each customer's carton represents an individual specification and that critical classification of color on one carton will not be critical on the copy styling of another carton. Careful considerations of equivalent values of monocolored lights at the customer's copy O. K. review, at the ink room, color lab, and at the press room . . . Munsel color standards . . . photovolt with Tristimulus filter excitation, etc. - these factors lend uniformity to the circumstances surrounding the evaluation. However, the key point is to spell out company policy. Is the control program a front office proposition that dies in the press room? . . .Or is it a program that stops the production of a below par quality job until quality corrections are made?

The obvious answer would appear to be the latter, but to be factual - this answer must be planned far ahead through quality education of sales people, proving department personnel, and production supervision.

a) The sales department must carefully check copy, colors, and style of a customer's product - where his sales experience dictates that this is a "rough" job to produce, he must be sure the production group understands the problems involved if a "tight" delivery date is required by the customer.

b) The proving room must carefully examine proofs, plates or engravings involved, bearing in mind the number of impressions involved . . . where questions arise that experience indicates a sharpness, flatness, shallowness, or poor fit may cause trouble, a decision should be firmly made to clearly spell out the quality evaluation so that production starts off with full knowledge of the problems involved. Many, many dollars can be saved by plate room planning in advance such typical measures as:

   "Die 14 is starting to become shallow - watch carefully after 15,000 impressions - replacement plate on hand."

   "Fourth color trap and register is tight - run through same press for second set of colors down - watch register marks and cycle closely to avoid sheet shrinkage."

c) The production department must learn that, percentage wise, a little bit of extra effort in fitting register will take care of most conditions of stock variation . . . that special spot packing on the cylinder will often help where surface of stock is rough and above all . . . that seldom can it be expected that a general make ready will fit stock as received.

With the three major contributors of variables, namely, sales, plate, and ink rooms and production geared to quality mindedness, the mechanics of press room control may be instituted. This may fall into two categories:

a) Where the level of material being printed is somewhat uniform, a rating is established for each defect by category . . . this, when evaluated with printing complexity plus surface area, yields a form of "C" chart. This technique has been pioneered by Sooey and successfully installed in a number of major carton plants.

b) Where "job" shop prevails so that quality runs from simple suit boxes to multicolor cosmetic cartons, the author has had excellent results with a bouncing ball chart. This technique is a two sigma variance chart that gives a running control with the inspector judging the printing complexity vs. type of defect. It is very sensitive to changes in process levels.

The conclusions that can be drawn is that three basic factors are necessary to build press room quality.

1. Strive for uniformity! Don't worry about the level of the process until you have licked the excessive variation.

2. Build quality mindedness so that everyone is working for the same goal - be sure people are sold as well as told a quality program is operating.

3. Win management's support, then keep management's support by earned respect that quality methods pay off. This cannot be done with words; it must be done with deeds.

Quality control can operate in the press room and will build better products through prompt corrective action because . . . when statistical charts indicate action is necessary, you can be sure that action is imperative, justifiable, and necessary!

QC CONCEPTS USEFUL IN OR

P. S. Olmstead
Bell Telephone Laboratories, Inc.

1.0 Introduction

Since World War II, Operations Research (OR) has received extensive publicity as "a scientific method of providing executive departments (and, therefore, executives) with a quantitative basis for decisions regarding the operations under their control".(1) This has excited the interest of many scientists and some of the more erudite executives. As a result, short courses similar in some respects to the 8-day courses in Statistical Quality Control, have been given at several universities. More recently, a few universities have organized Operations Research curricula. This development has been in the direction of establishing Operations Research as a new science.

However, others, particularly engineers, have been inclined to consider OR primarily as the application of recently developed mathematical techniques to the analysis of complex situations. As such, it is a part of Industrial Engineering, Systems Engineering, and, of course, Quality Control. The fact that it may be a part of each of these and many others suggests that people from these several fields may wish to get together to discuss their diverse problems. This should obviously lead to more wide-spread use of specific techniques that might otherwise be developed for use in only one field. A simple example is the extension and application of Quality Control (QC) techniques to clerical work, to merchandising, and to many other fields.

It is just as important for QC people to find out more about OR as it is to find out about the latest development with the particular product for which they are responsible for quality. At the same time, it is important for QC people to inform OR people about fundamental theory of QC that should be useful in applying OR to other fields. That is the primary objective of this paper.

An OR problem that each QC person has encountered is the determination of an assignable cause. Ordinarily, this would be taken to be simply QC. However, it is a problem that might be assigned to an OR man to solve. But will his method of solution be different?

2.0 Basic OR Process Applied to a Specific Problem

If a person is assigned to solve a particular problem, he must first be told about the problem, i.e., what is wanted. This is not always easy. Sometimes the want is contingent on what is found during an investigation. For purposes of simplicity, it may be assumed that the problem here can be reduced to finding the cause of some undesired effect that has been observed in a process. This may be formalized as

2.1 Statement of Problem: To find the assignable cause associated with the undesired effect observed in the process.

The obvious next step is to take a look at the process. This may involve reviewing the theory relating to the operation, the troubles that had been encountered previously, and the data that are currently being obtained. In some cases, it may even be necessary to obtain new data. At this point, it may be worth while to mention something about how this is done in OR. Where ordinary scientific method might assume this to be a one-man job, OR recognizes that most operations are complex and may require consultation with a number of people. Usually, the assignment for solving a problem is given to a team rather than to an individual. The leader of such a team then acts primarily as a coordinator. Formal designation of this step might be

2.2 Collection of Information for analysis from all available pertinent sources, including

1. Accepted Theory for normal operation from scientists or designers,
2. Possible Hypotheses about troubles from scientists or engineers, and
3. Data, both existing and new, from QC or other records.

The emphasis in the OR approach is on people and the ideas that people have. It is recognized that the same set of data mean different things to different people. For this reason, a group approach to a problem is apt to give a better answer than that obtained by one man working alone.

In trying to find the assignable cause associated with an undesired effect observed in a process, it is likely that the OR man would first determine the intent of the process and the background of its development from the engineer who designed it. This is to provide a picture of accepted Theory relating to the process. It also may disclose information relating to troubles encountered during development, together with their effect on the output of the process. This may suggest possible Hypotheses about the assignable cause being sought. The next step is to see what is actually happening in the process. On the basis of accepted Theory and the possible Hypotheses, existing Data are examined, the process reviewed with the engineer in charge, and in some cases new Data obtained in order to secure a more complete record of associated happenings. Sometimes, the process review suggests new possible Hypotheses which require additional new Data. The extent to which the search for Information would be continued depends on the experience of the OR man and his judgment about the sufficiency of the Theory, Hypotheses, and Data for the next step in solving his problem. This step is a

2.3 Critical Analysis of Information

In many respects, this step in an OR investigation is applied statistics. In the statement of the problem, it was assumed that originally the process was free from the assignable cause now believed to be present. This can be tested by showing that before the present situation developed the process was in control. Control chart presentation of past Data, therefore, should show that control had been attained and thus confirm the reasonableness of accepted Theory. Continuing the control charts to recent Data should show that at least part of the Data is outside control limits. This would confirm the assumption on which the investigation was started, namely, that an assignable cause is present. To find out what the cause may be requires other information.

A starting point is to consider each of the Hypotheses and the new or old Data that relate to them. In some cases, Design of Experiment may have provided Data suitable for an Analysis of Variance that is carried out. In others, particularly where QC charts have been in use, results of investigations of out-of-limit performance may be available. In general, these would be expected to show that change in practice had occurred at the time of the poor performance. Any one of several possible ways might be used to identify the assignable cause sought.

Having located the assignable cause through such critical Analysis of Information, consideration is given to preparing a

2.4 Predictive Statement that summarizes

1. Limitations of existing Theory,
2. Necessary new Hypotheses, and
3. What new Data may be expected to show.

In the problem relating to the found assignable cause, this Predictive Statement might take the following form:

A. After taking steps to eliminate the found assignable cause, it will be possible to attain control of the process consistent with existing Theory, but

B. On the Hypothesis that the found assignable cause will continue to affect the process, control of the process consistent with existing Theory cannot be attained without process modification.

In some cases, a more specific statement relating to the expected average and standard deviation of the quality characteristic involved is possible. Many OR analyses set forth the cost factors relating to various alternatives such as A and B but this need not be considered basic to the OR method.

3.0 QC Comments on the Basic OR Process

Many QC people may feel that the basic process outlined differs but little from what would be the corresponding basic QC process. However, this apparent similarity has been enhanced by citing an example from the field of QC. Most OR problems do not have QC backgrounds as OR people might view them. The usual QC approach is to assume that each problem is one of a set whose desired end result is a state of statistical quality control. The problem changes slightly with each repetition but inevitably leads to what is acceptable as an economic solution. From the viewpoint of QC, each of the four steps in the basic process is apt to be a cause of trouble or assignable cause. Experience in solving other QC problems has indicated the nature of some of these.

It is generally recognized that the Statement of a Problem is often incomplete. Modifications may have to be made to take account of costs, differences in what individuals or groups of individuals want, changes in what they want, and even the availability and the properties of materials. Even in the simple problem of finding the assignable cause, there are the implied problems of what can be done about it and what level of quality is to be expected if the answer is nothing. This omission from a complete statement of the problem might not become apparent until an attempt is made to make a prediction.

It is one of the fundamental assumptions of QC that Information is never complete and that such Information as is available may be misleading. One possible reason is that the individual or individuals responsible for collecting it may overlook important evidence; another, that accepted Theory is sometimes false. This may lead to omitting the right Hypothesis as one of those to be investigated. Then, too, some Data are unsatisfactory as evidence. Another of the fundamental assumptions of QC is that effects of undetected assignable causes are usually present in every set of Data. If these assignable causes cannot be detected by applying QC procedures, the Data are unsatisfactory.

When the Information is not complete (see 4.3), it is almost certain that the Analysis will be unsatisfactory even if it makes maximum use of the available Information. However, even when the Information is complete, the Analysis may still be unsatisfactory. This will be true if no attempt is made to detect and find assignable causes of variation. Often this occurs because the analyst assumes that the Data constitute a random sample without making a statistical test to see whether or not this may be true. Determination of expected variability on the basis of accepted Theory for comparison with actual QC performance is sometimes omitted. Again, actual performance may be accepted erroneously as the limit attainable in the process.

Each of the foregoing contributes to the validity of the Prediction that can be made. To be good, a Prediction must be based on a good Statement of the Problem, good Information, and good Analysis. A Prediction is apt to be poor if accepted Theory is inadequate or incomplete, or if accepted Theory is insufficient for the practical problem. A complete Prediction must not only show what must be controlled but also what will happen when control of each important component is not attained. Some Predictions are good insofar as they indicate the expected average of a process but poor in indicating what may be expected for individual values. This can be avoided by using the QC concept of tolerance range. Predictions relating to tolerance ranges for individual values and for averages have been demonstrated in QC as most informative and useful for setting up schemes for continuing good performance in a process.

In these comments on the basic OR process, reference has been made to various QC assumptions that have been verified as being true in almost if not all applications of QC methods. These assumptions relate to experience with manufacturing a product. Conceptually, however, they would appear to apply with equal validity to any repetitive process. The next section discusses some of these concepts that can be applied to the OR process.

4.0 QC Concepts Worth Applying to the OR Process

From the viewpoint of QC, all processes are repetitive in nature. This is particularly apparent in all stages of mass production. It may seem less apparent in other fields of application. However, if it were not for the repetitive nature of signing a check, people might have more difficulty than they do about having them cashed. The authenticity of a painting as judged by an expert depends on many items of a repetitive nature. It is not unreasonable, therefore, to expect to find elements that stay constant or change gradually in everything even in processes of thought. On this basis, the QC approach to the OR process assumes the concept that

4.1 The entire OR process is one cycle of a Repetitive Operation.

In setting down the steps of the OR process, no provision was made for such a repetitive operation. However, in discussing the steps, it became quite apparent that when something new was learned by taking the step, it often suggested that something should be changed in a preceding step. This is indicative of the cyclical nature of the process. The steps cannot be considered to be independent.

Viewed in this way, each so-called new problem becomes simply an up-to-date statement of an old problem. Solution of any problem is simply an economic stopping point. With tomorrow's data, the available information will be changed and the solution may be different. Today's answer, to be complete, must account for each of tomorrow's possible conditions as they are anticipated today. Often, today's answer is like yesterday's except for normal sampling variation or some small cumulative deviation that may turn out to be a trend. When today's answer differs in an important way from yesterday's, it is time to look for an assignable cause. This may have been introduced by a new statement of the problem, by introducing new information, by using a new method of analysis, or by expressing the prediction in a new way. The OR approach that emphasizes group examination of problems is very apt to turn up something new so that today's answer will be different from yesterday's. However, the OR answer to be acceptable must tie together yesterday's answer with tomorrow's possible answer as part of a QC type repetitive operation.

Accepting this concept of repetitiveness, it is a simple extension to accept the concept that the OR process is one cycle of a repetitive operation. Then, if the entire process is repetitive, each element in

it is also repetitive. Of course, this means that change in any important element is apt to result in a change of each of the other elements. This is known to occur in many, if not all, QC situations and may be presumed for each OR situation. The desired end result of QC considerations is a state of statistical control. It is reasonable that this should also be the end result desired in an OR investigation because predictions based on such a state will have maximum validity. One of the important problems of QC is the establishment of a process for reaching such a state. For mass production, the major features of this process are well known. The process is less well defined for determining QC with respect to consumer wants, research and development, design and specification, and field performance. To the extent to which QC and OR overlap in these applications, it will be worth while to consider those QC concepts that have a bearing on the steps of the OR process.

4.2 QC Concepts Related to Statement of the Problem.

It has been pointed out that the statement of the problem is seldom complete. It is the prospective consumer of the results of the process who phrases the original statement. It is his statement of what he believes he wants. It is based on the results of previous allied problems and is subject to modification when the initial results of this investigation become available. QC Theory suggests that an adequate statement of the problem will be impossible until the final results of the problem are known and have been taken into account. This is the equivalent of reaching a state of statistical control, a specific concept of QC.

In some cases, the statement of the problem does not take into account the ultimate use that the prospective consumer has in mind. If there are several prospective users, each with his own application, an all inclusive statement of the problem may not even be possible. In this situation, one choice is to reduce the problem to an average that may be sufficiently close to satisfy most of the users. However, to the extent that the wants of the users can be divided into subgroups, it is sometimes possible to make separate solutions for each. In QC, this is division into homogeneous subgroups or by assignable causes, another specific concept of QC.

4.3 QC Concepts related to Collection of Information

The interrelationship among Theory, Hypotheses, and Data has been considered at some length in QC literature.(2,3) It is recognized in QC that not all data are good. Some may come from a process that is in a state of statistical control but others may not. In most cases, experience has shown that they do not. For this reason, QC has introduced the concept of an Assignable Cause and insists that all Data should include sufficient information to permit detection and identification of such a cause. On this basis, all good data should include:

1. The Order, i, associated with each measurement,
2. The Condition, C_i, under which each measurement was taken,
3. The Observer, H_i, the human being who made the measurement,
4. The Time, t_i, when each measurement was made, and, of course,
5. The Numerical Value, x_i, of each measurement.

This implies that, in a theoretically exact experiment, it should be possible to determine an equation:

$$x_i = f(C_i, H_i, t_i)$$

Actual experiment, however, has never been sufficiently precise to permit such determination as a general rule. Changes in conditions are not determined exactly. Even a single observer cannot be considered as a constant. For these and similar reasons, the best that can be done is to examine the repetitive series, x_i, for statistical control and use the available information, C_i, H_i, and t_i, as clues concerning the Assignable Causes that may be detected.

In order to be able to make valid predictions, data from a process must give evidence of statistical control. This leads inevitably to the question, How many observations are required? As reported in QC literature,[2] experience has shown that a sequence of at least 25 samples of four that satisfy a specified criterion for control is required to give even a preliminary basis for judging that a state of statistical control exists. To establish a satisfactory Tolerance Range, another QC concept, for individual measurements requires even larger sample sizes.[3,4] Thus, on the basis of QC considerations, it may be concluded that

A. Less than 100 measurements may be sufficient to disprove a <u>Theory</u> or <u>Hypothesis</u>, but

B. At least 100 measurements that meet a specified criterion for control are required to provide a minimum satisfactory basis for <u>Prediction</u> relating to average and standard deviation.

4.4 QC Concepts related to <u>Analysis of Information</u>.

The concept of Statistical Control takes precedence above all others in any QC Analysis of Information. This assumes that a criterion of control has been established as part of accepted Theory. Depending on the problem, the criterion may be any one of several possible criteria. It may even be a combination of several. In inspection work, it has been customary to rely on the use of control charts. In other applications of QC, many other criteria are used. The choice depends on the economic factors of cost and the value of finding an Assignable Cause. This QC concept, Assignable Cause, has been defined[2] as a cause "that can be found by experiment without costing more than it is worth to find it". It is obvious that many causes that would be considered assignable when determining a fundamental constant of physics or chemistry would not be so considered in controlling a manufacturing process or even in carrying out some researches or developments.

Not all possible Assignable Causes are known at the start of an experiment. For this reason, a suggested QC procedure is to apply a number of statistical tests to the Data to detect any unknown Assignable Cause and to predict its nature.[3] This is also proposed as an important tool of analysis that should be applied to OR situations. A discussion of the details of such application is outside the range of this paper. However, several pertinent articles exist in QC literature.[3,5,6,7]

QC has always recognized that control of quality is a cooperative venture. In production work, this has meant cooperation between inspectors and engineers responsible for making the product. Even the engineer responsible for the design must be included in the team. When the quality to be controlled is in research and development work, the team may include engineers, scientists, and statisticians. Depending on the problem, these scientists may include psychologists, econometricians, biometricians, and so on, as well as physicists and chemists. Each individual in such a team brings a knowledge of accepted theory in his own field that may help to account for Assignable Causes detected as noted above. In many respects, this joint or Cooperative Investigation by engineers, scientists, and statisticians is similar to what is intended in an OR investigation.

However, there is a difference in what happens at the end of the investigation. In the case of OR, the problem is often considered completed as soon as the analysis is complete and the report written or made. This is not true of QC. Finding trouble requires that action be taken. Usually, this action is initiated by the QC man himself. This means that he has a responsibility to see that he has reached the right conclusion and that the action does in fact improve quality. If it does

not, it is necessary to keep looking for the cause of trouble. Thus, the inclusion of Corrective Action in the QC procedure is an important governor that has assisted in sharpening the tools of QC and might be equally useful in OR.

4.5 QC Concepts Related to Prediction

The solution of a problem is necessarily a Prediction relating to it. QC has given considerable attention to the reliability of Predictions. One conclusion relates to the kind of information on which the Prediction is based. Stated formally:

A. A Prediction has maximum validity only when all Assignable Causes relating to it are known and have been taken into account.

Thus, a Prediction is meaningful in a statistical sense only when it applies to a situation that can be demonstrated to be in a state of statistical control. Since this is a condition that may not be met. QC has considered alternate situations as well.

In the design of a piece part to be used in a product, the engineer carefully specifies limits to be met by each dimension. A critical instance of this type is the design of a shaft to be used in a bearing. Ordinarily, a sloppy fit is just as important to avoid as an interference fit. Thus, from the point of view of utility, there is a desired clearance on each side of which value decreases. The engineer sets tolerance limits for individual shafts in order to limit the number of interference and sloppy fits in the assembled product. The same sort of consideration applies to the bearing as well as to the shaft. At first, it may seem that knowledge of the average and standard deviation of shaft diameters should be sufficient to assure a satisfactory product. Unfortunately, this is not so. Even if out-of-roundness is excluded, wear of tools produces nonuniform variation in the product that must be considered. For this reason, the practical average of the engineer is at best a range of averages and not a true average. This discussion leads us inevitably to the following conclusions:

B. That Prediction relating to tolerance limits or a Tolerance Range for possible individual measurements is equally as important as Prediction relating to the location of the Average, and

C. That Prediction relating to the location of the Average should also be in terms of a Tolerance Range for an average of at most 100.

The second of these conclusions is, in effect, a limitation on what is considered an Assignable Cause. It recognizes that a shift in the average equivalent to one twentieth of the range between the tolerance limits will have little effect on the number of individuals found outside these limits and also would require an unduly large sample of observations to verify.

Another QC observation relating to Predictions is that attention should be directed to the control of Assignable Causes. It is for this reason that inspectors are furnished lists of defects relating to the products that they are inspecting. These lists of defects are not static things that may be determined and then forgotten. Instead, they are simply the latest version of the troubles that are being encountered. Inspectors are expected to add items to these lists as new troubles are encountered and to omit items for which technology has found adequate answers. This leads to the conclusion:

D. That the Prediction should make clear what independent checks should be made to assure that future measurements will have

been taken under conditions appropriate to the Prediction, i.e., that known Assignable Causes have been brought under control.

5.0 Summary

This paper assumes that both OR and QC employ scientific method. The aim of both is to establish an economic answer. To date, QC has had its widest application in the field of manufacture or, more generally, in the field of repetitive measurement. In a restrictive sense, OR has also been applied in the field of repetitive measurement. However, where QC has usually been used to reduce variability in a given process, OR has been used to compare several processes. Both are interested in cost. In QC, it is reducing cost by eliminating assignable causes, usually in a single process; in OR, it is comparison of the costs of several processes Both, to be well done, require the cooperation of engineers and scientists of all types.

A major difference between OR and QC is that OR seldom assumes responsibility for action while QC assumes that initiation of action is part of the job and is to be followed by monitoring the results of the action. This has helped QC to obtain a rather complete picture of the repetitive process that should be helpful to OR people. This paper draws attention to certain concepts used in that picture that have possible applications in OR.

It is fundamental to QC that the solution of all problems involves a repetitive process. Even the statement of a problem is one step in such a process. Theory, Hypothesis, and Data change from day to day. Today's analysis differs from yesterday's. Inevitably, predictions should become more meaningful with each repetition.

Specific concepts developed from QC considerations include:

1. Components of data as evidence
2. Statistical Control
3. Assignable Cause
4. Tolerance Range
5. Components of prediction with maximum validity

QC has found that each of these concepts is necessary for a complete evaluation of the repetitive process. It seems reasonable, therefore, that they should be equally important in OR considerations.

References:

(1) Methods of Operations Research, P. M. Morse and G. E. Kimball, Wiley, New York, 1951

(2) Statistical Method from the Viewpoint of Quality Control, W. A. Shewhart, the Graduate School, The Department of Agriculture, Washington, D.C., 1939

(3) How to Detect the Type of an Assignable Cause, P. S. Olmstead, Industrial Quality Control, Vol. 9, Nov. 1952 and Jan. 1953

(4) Tolerance Limits for Normal Distributions, A. H. Bowker, Chapter 2 of "Selected Techniques of Statistical Analysis", McGraw-Hill, 1947

(5) Textile Yarn Abrasion Test, A. C. Walker and P. S. Olmstead, Textile Research Journal, Vol.15, pp 201-222, 1945

(6) Tracking Troubles in Atmospheric Corrosion Testing, P. S. Olmstead, W. E. Campbell, and H. G. Romig, ASTM Symposium on Atmospheric Exposure Tests on Nonferrous Metals, pp 76-88, 1946

(7) Principles of Sampling in the Chemical Field, M. A. Brumbaugh, Industrial Quality Control, Vol. 10, Jan. 1954

THE APPLICATION OF QUALITY CONTROL
TECHNIQUES IN DETERMINING WORK ASSIGNMENTS AND STANDARDS

Solomon Barkin, Research Director,
Textile Workers Union of America, CIO

Quality control techniques are being gradually introduced into the textile industry where operating controls are greatly needed. The existence of a multitude of interdependent variables has always made for highly unstable operating conditions in textile plants. But few sustained efforts had been made to stabilize them although many managements have sought to improve the general level of operations. Attempts to set work assignments, and to control the quality products have brought managements to grips with the problems of variations. Yet the time study men responsible for setting job assignments have tended to dodge these problems and have avoided quality control techniques. Production managers have also taken negative attitudes and continued skeptically to insist on proof that controls pay off.

The time study men's failure to use quality control techniques for defining physical phenomena has created a gap between their claims for, and the results of, their work. While they claim to have been fixing stable equivalent levels of work application on all jobs, paying premiums for work above such levels, few have attempted precisely to define the level of expected performance of each quality precisely enough so that the actual level of the work demands could be ascertained. As a consequence, one necessary basic, though not the sole, change in time study theory and practice must be the adoption of quality control techniques to describe more accurately the jobs which these men construct for employees.(1)

Workers, like buyers of merchandise, have understood that they must protect themselves from vague definitions; they protect themselves from poorly defined job standards by demanding lower work assignments. Many serious industrial disputes arise from management's unwillingness to guarantee a specific set of physical conditions and the workers' desire to avoid excessive physical demands during periods of extreme job deterioration.

Quality grades in the merchandising of textiles and other materials have been introduced to eliminate speculation over the likely quality level of the goods being bought. The buyer of ungraded fabrics usually offers a lower than prevailing price for them to hedge against a higher than customary percentage of low grade fabrics. The absence of careful definitions of grades makes proof of quality difficult and the haggling over price has, therefore, to consider not only the economic value of the specific quality, but also the risks taken on the actual quality of the merchandise. The establishment of specific product grades has therefore been advantageous both to the seller and the buyer. Premiums paid for high quality have acted as an incentive to improvements. Grading has eliminated a risk from the buying process. The purchaser no longer has to hedge his price against this factor. Workers would also moderate their attitude toward many industrial problems if they were offered carefully defined statements of work conditions which were to be strictly maintained on a job.

To us in the textile industry where erratic work conditions are particularly prevalent, the establishment of controls is of prime importance in any procedure for work measurement. The quality and processing characteristics of the raw materials used in each process affect productive efficiency. The machinery is almost wholly automatic, continuously processing raw material until it either exhausts its supply reservoir, fills its delivery reservoir, or until material breaks or mechanical difficulties arise. Breaks may result from a number of causes, such as the characteristics of the original raw material, deviations induced by the prior processes, machine and operating difficulties, or worker handling. The worker's task is primarily that of loading supply packages for processing, removing the fully produced package or doffing, maintaining the machine and the cleanliness and orderliness of the room, and repairing broken strands in the material as it is processed to restore machine activity. Most significantly, the demands made on workers by any specific battery of machines will vary with changes in these factors and the frequency with which any specific service has to be performed. While some frequencies can be predetermined by management through scheduling, others (periodic duties) are related to the production cycle and the machine's efficiency is controlling. For still other frequencies (random duties) estimates can, at best, only be based on statistical studies of past experience.(2) But with the great likelihood of changes occurring in conditions, these data may become unrepresentative for future operations.

While a textile worker's assignment is characteristically rigid in terms of a specific stand of machines, the demands on him may vary widely from hour to hour and day to day. The very predominance of random work duties in job assignments makes for great uncertainties as to actual work requirements. Very often the work duties accumulate and the worker is simultaneously beset by many demands for service. A successful training program and good morale will instill that pride of job in workers which will makes them intolerant of idle machines and drive them to restore the machines to production, thereby creating unusually high demands on them. But the continuance of such high pressures will tend to demoralize workers, create great dissatisfaction, and may also produce defensive indifference, and possibly, retaliation, a not unusual by-product of the time study man's activity in the textile industry. The destruction of the pride of craftsmanship through excessive work demands resulting from unstabilized job demands has also resulted in progressively poorer products and growing production difficulties.

While there are few industries where the merger of the techniques of quality control and time study is as essential as in textile, the latter is one in which the use of quality control procedures has been extended only slowly. Its full integration into the production and managerial processes is rare.

Few time study practitioners have undertaken a systematic description of the physical realities of an operation in the terms employed by the quality control analyst.

We in the trade union movement have been particularly conscious of the absence of such integration and have urged, through a publication entitled "The Textile Workers' Job Primer," more extensive use of quality control approaches as a prerequisite to more intelligent bargaining

on the size of the work task.(3) We want the worker, like the buyer of merchandise, to be relieved of the risks of uncertainties in the definitions of his job. The buyer has been guaranteed the quality of merchandise; the worker wants his task defined in terms of carefully itemized job demands with specific notations on the frequencies of duties to be strictly maintained within a prescribed range of variation.

SYSTEMS OF CONTROL IN THE TEXTILE INDUSTRY

Work assignments must necessarily be related to the degree of stability attained on a specific job. To define stability, we must have tools for measurement and methods of recording the occurrences. To establish confidence in these standards, there must be a system of control which insures detection of departures and immediate correction of conditions causing such undesired variations. It is well to review the prevailing procedures in the textile industry, for then we shall be better equipped to determined the method of setting job standards best adapted to this industry.

CORRECTION SYSTEM

In any industry where manufacturing is still largely an art, and scientific information is limited, the primary reliance for operations must still be placed on the personal experience of the supervision. The production men are expected to follow the manufacturing operations, detect deviations, and correct them. Depending upon the individual's acumen and the conditions which develop, extreme failures are likely to be detected and efforts made to remedy them, particularly if they persist for a noticeable period. Special attention is likely to be paid to the characteristics of the final product, especially in times of a buyer's market, when products may be returned by the customer.

The procedure emphasizes correction of extreme departures from customary levels of variation, individually interpreted by supervisors. Correctives are applied with the hope that they, or other countervailing forces, will rectify the condition and restore the customary levels of variation. But actually, the range is not known or defined, and can vary widely from person to person, machine to machine and from time to time.

In such plants managements, in the past, may have tried to shift the onus of bad results to workers, by imposing fines, partly with the hope of thereby enlisting the workers' efforts in the detection and correction of bad conditions. Unfortunately, workers in most cases could not identify the trouble and frequently found their suggestions unwanted and rejected. Under those circumstances, workers have not only had to work harder to overcome defective conditions, but have suffered drops in earnings from reduced machine efficiency and also have had pay deducted for the lower quality output. Fortunately, union pressure and a higher sense of managerial responsibility have combined to eliminate these penalties.

Workers have responded to these conditions with various defensive practices. A common one is to demand an organized system of preference or equal distribution of jobs to assure either better treatment for the senior employees or equal distribution of jobs among the employees.

Another reaction has been to insist upon low enough assignments so the job will remain tolerable even in face of the most serious deterioration of operating conditions.

The correction systems face their severest tests during periods of intense competition when management seeks to extend assignments to meet competitive costs, but refuses concurrently to guarantee job conditions. Plants operating with such systems cannot provide responsible calculations for work assignments.

PREVENTION SYSTEMS

Many mills have advanced beyond the above system by instituting standards for the purchase of raw materials or for the operation of specific parts of their manufacturing processes, such as room and machine conditions and settings. Management, either on the basis of specific studies, or its own experience or that of others, determines the conditions and settings likely to yield the best operating results and the desired product quality. These then are followed in the purchase of the materials, the operation of the mill and the location of production facilities.

A prevention system represents a marked advanced in textile operations, since it defines management's responsibilities for supplying specific materials, machines and surroundings. Departures from standards set in these areas may be carefully evaluated and brought into line. Each specification can be tested for its appropriateness and revised as knowledge improves or experience dictates. Some mills have not only set up such specifications and followed them in practice, but have also instituted careful systems for checking the degree of conformance of actual conditions to these standards. The most advanced development in this area is the institution of a preventive maintenance program providing for the regular examination of machine parts and conditions and replacement of machine parts not meeting the specification, even though they have not affected production. This preventative attitude contrasts strikingly with the emphasis on simple correction previously outlined.

While the above practices are likely to reduce variations, they do not provide a definition of the range of permissible variations or a careful program for evaluating or controlling them. They begin with the assumption that if conditions are properly set, an absolutely stable population of variable values can be maintained on a long-time basis. This position is unrealistic, particularly in the textile industry. Each fibre is unique and its processing characteristics cannot be easily predicted since they will fluctuate continuously. Machine and room conditions are constantly changing under the influence of such factors as handling and atmospheric conditions. Full implementation of this preventive technique would require precise knowledge of the effects of each of the possible values of the other variables and their myriad of combinations. Such knowledge is unavailable and is not likely to be manageable. Invariable conditions cannot be maintained.

This prevention, therefore, neglects to treat variations in actual results. It uses averages as standards without necessarily taking into consideration the fact that two equivalent averages may represent completely different basic distributions of values with consequent diverse

demands upon workers. The supervisory official using such specifications often becomes so imbued with a feeling of their appropriateness and infallibility as to be completely deaf to considerations of the meaning and effect of variation. Managements following this technique often uncritically lend themselves to the use of the current time study practice, without appreciating its failings as a measuring tool in a textile mill. Work assignments are likely to be set more tightly and rigidly in such mills, with few concessions to the actual variations in conditions developing on the job. Only in the case of extreme departures from customary operations, recognized either by the worker or the management, will corrective steps be taken.

CHECKING SYSTEMS

A number of companies have made a further advance by instituting systems for checking both the frequency of random occurrences such as yarn breaks and loom stoppages, and other variable operating conditions like bobbin weights, machine speeds, package weights and product characteristics such as neps and strengths. The objective is to test conformance of the average of these qualities with a standard rather than to measure the range and nature of the deviations.

These tests for conformance to standards are seldom precisely defined. First, samples are not systematically obtained; the sufficiency of their number is not determined; the adequacy of the size is not evaluated. Second, the degree of conformance is seldom, if ever, prescribed. Third, the range and nature of the dispersion are not regulated. The checks tend to be sporadic.

The procedure flows from concepts existing in the prevention system which seeks to enforce given specifications. The aim is to approximate a standard rather than to assure stability.

QUALITY CONTROL TECHNIQUES

Unlike the preceding, the quality control techniques utilize not only specifications for purchasing the proper materials, setting the appropriate conditions and checking on operating conditions, but also prescribe procedures for checking deviations from standard operating conditions, product qualities and results. The quality control man, in collaboration with the manufacturing engineers, supervisors and the laboratory technologists, determines objectively the extent to which prevailing variations in both mean values and dispersions about means can be attributed to chance or to specific causes. Specific disturbances creating unusually erratic behavior can be isolated and corrected. After they are eliminated, the quality control techniques are used to maintain specific levels of performance and ranges of variations, and help in the improvement of the level and the narrowing of the range through careful study of the information obtained on the control charts. Most important, the quality control system presupposes continuous, permanent and representative checks on performance and therefore reflects management's acceptance of its responsibility for maintaining standard and stable conditions.

A quality control program for a textile mill must include (1) the study of stock characteristics to secure stable mixes and provide

immediate guidance to the operating staff on the nature of the changes in stock mixes as they develop, particularly in respect to the natural fibres, but to some extent also with synthetic fibres; (2) the checking of room and machine conditions for conformance with permissible ranges of variation or for replacement; (3) the checking of actual machine speeds and settings as well as parts, for adherence to specifications and, where variations are permissible, for observance of the control limits on variations; (4) sizing of the products at each step in the process, to determine whether the weights are stable and conform to specifications; (5) the control of the degree of linear regularity or evenness of the product within prescribed limits and other product qualities.

In addition to the above program of control of conditions and products, it is also necessary to have specific information on factors directly affecting the work demand. Averages and ranges of package weights, for example, directly affect the frequency of some periodic work duties; and the frequencies of material breaks or machine stoppages materially determine the work required of the attending employees. The measures must, of course, be directed not only to checking the mean frequency, but also the variations in the frequency, thereby permitting determination of the degree of conformance with the control limits.

The institution of a complete system of quality control provides a realistic basis for testing the level of performance and also the degree of observance of the control limits. It is the only control system which provides specific control limits. As such, it offers a definitive basis for calculating work assignments. The other approaches to control base their estimates on standard frequencies ignoring the effects of the higher frequencies. The use of the quality control technique makes it possible to set a ceiling on work demands and to use this ceiling as the critical determinant in the formulation of work assignments.

This approach - using upper control limits rather than average levels of job demands - has been recognized as basic to the realistic application of time study technique by arbitrators. Practitioners have, unfortunately, continued to ignore these findings. One decision, in the automobile industry, held that the job level which the management can claim as standard is the upper limit above which the worker cannot be expected to apply himself. Management could not, the arbitrator held, justify temporary operation of the assembly line at above standard speed merely because the line had been stopped part of the previous hour and it wished to maintain the established average output for the shift.(4) This conclusion has been reinforced by other arbitrators dealing with piece rate jobs in other industries. In textiles, arbitrators have held that the workers' negotiated hourly rate is exactly what the term implies, a minimum hourly floor. Bonus earnings accruing in any single hour may not be used to make up pay because the worker fails to earn his guaranteed minimum during another hour. Each hour must be tested separately.(5)

The trade union movement has, therefore, increasingly favored the use of the quality control technique for the definition of work demands, because it presents the limits of permissible variations (usually referred to as control limits). It is a technique, acceptable to both advanced management and workers, for defining measures and limits of frequency. It provides a scientific guide for the operation of sampling

procedures to secure representative results, and control charts to inform management automatically when conditions deteriorate to the point where control limits are exceeded and corrections are imperative. With the use of quality control techniques, management is adopting a procedure for calculating job assignments in terms of actual rather than supposed job requirements and in a form which automatically carries with it the responsibility for maintaining these conditions.

ORGANIZING A MILL FOR QUALITY CONTROL

The institution of a complete system of quality control represents a radical alteration in managerial methods. It constitutes the substitution of definitive measurement for rule of thumb hunches. Evidence, rather than mere hearsay, is more likely to be used in such programs in the development of specifications, standards and control limits. Secondly, it means the systematic recording and formalization of production practices. Thirdly, it results in the close coordination of the laboratory, testing officials, production supervisors and management. This latter objective will be most difficult to attain and many a program has floundered on this rock. To operate the program effectively the management must not only accept the theory of control and have the conviction that it is economically sound, but it must also be competent to handle the techniques and master the meaning of the charts and results. It must also have the determination to accept and act on the warnings sounded and the conclusions suggested by these charts. Policing the control limits and enforcing their meaning becomes critical both for the operation of the mill and the worker. unless these control limits are used strictly and automatically, the entire procedure deteriorates and becomes useless as an approach to the determination of work assignments.

There are few mills in the textile industry which have a complete integrated system of quality control to regulate stock, machinery, settings, sizing, linear regularity, package sizes, product qualities and random element frequencies. Some progress has been made, particularly in the measurement of raw materials, sizing and linear regularity. But even with respect to these items, few maintain complete control systems.

Too many managements assume that, in view of our lack of knowledge of many phases of the industry, wide variations and instability are inherent. Only recently, as new measuring instruments have been developed, have specific tests become frequent. But they have seldom been integrated into a complete control system. Most managements have resisted the costs and remained skeptical of their relevance to the determination of work assignments or to the lowering of manufacturing costs.

THE TIME STUDY MAN'S VIEWS

Before evaluating the effects of the preceding discussion upon the methods of determining work assignments, it is necessary to describe the time study man's approaches. His failure to assimilate the significance of quality control techniques for his problems stands in the way of reducing this area of controversy between management and labor and developing a satisfactory procedure for calculating work assignments.

The fundamental inclination of the time study man is to eliminate all job complexities from his calculation so that they may be easily made. Variables in frequencies are dismissed by the use of averages obtained through haphazard and unchecked sampling techniques. He blinds himself to the fact that an average is, at best, a summary description of a distribution of values and is hardly an adequate description of the entire distribution. He makes the convenient assumption that the selected figures would be an invariant absolute. His glib answer to those who point out that deviations occur constantly is that they will be averaged out, hoping his response will not be tested by a trade union before a careful arbitrator or responsible and informed management.

The fundamental data used by time study men to obtain their representative figures are seldom built upon careful sampling procedures. Usually they are derived from sporadically conducted observations of varying duration. Sometimes, tests are made to satisfy a management's curiosity or a worker's grievance or because observers are available. Seldom are they made on a thorough and continuous basis. The acme of scientific inquiry usually is represented by the pretentious concession that the results should be built on full shift counts of highly variable frequencies such as end breaks or other random occurrences. But no such test can be representative of any conditions except those being studied, and then only if the range and character of the dispersion are also noted. Time study men often contend that the procedure of making eight-hour studies is "better than sampling." They fail to realize that the eight-hour studies are merely samples for an infinite population of shifts. The fundamental fact is that they are unwilling to acknowledge the presence of uncertainties associated with all sampling and the need for defining the extent of the variations. They cling to the concept that they must deal with fixed absolute average values. They have no place within their system for the concept of confidence limits.

Happily, arbitrators are beginning to enforce a more realistic approach upon time study practitioners. They are beginning to accede to union requests for a more thorough analysis of the data underlying time study summaries, which is proving most enlightening. Arithmetic means built upon a very limited number of observations encompassing wide ranges of variation have been rejected. Data purporting to be representative, when the probability of the observed mean being within plus or minus five per cent of the true population is not very high, have been held to be meaningless.(6)

While these awards are encouraging with reference to their effect on the sampling procedures themselves, they have not inculcated time study practitioners with the importance of tailoring standards to the type of data employed to build them. Appreciation of the pronounced difference between variables and absolute constants has never been grasped by most practitioners.

The assumption of absolutes leads the time study man to believe that a job standard is a stable fact, requiring little policing or review. Despite the protestation that time studies are necessary to insure a correspondence between pay and application, they abandon the strict enforcement of this conviction in the administration of standards. There is no procedure for self-administration of a standard to insure

that changes in labor requirements will bring an automatic correction in values. Changing frequencies do not immediately result in new values, for that represents too complex a challenge for current time study men to deal with. They would need an elaborate system of control over work duty frequencies, as provided by quality control engineers, which they are not prepared to insist upon, as a necessary prerequisite to their own operations. They prefer rather to dismiss the entire problem with the consoling belief that an average is a representative figure which irons out differences. Various allowances are added to the job standard to take care of the unusual factors.

These views offered in defense of the time study man's practice have, of course, made him the object of increasing attacks on many grounds. Many of us have wondered how the practice has been able to maintain its hold in face of these onslaughts.(7)

THE USE OF QUALITY CONTROL TECHNIQUES IN WORK ASSIGNMENT DETERMINATION

The current procedures for setting job standards and assignments require substantial overhauling. Not the least significant change is the one needed to assure a closer relation of the assignment to the level of actual physical stability prevailing on a job. The level and dispersion of the frequency of work duties must be clearly known before assignments are made. Certainly, there must be a procedure for correcting assignments and work standards when they seriously depart from assumed levels.

To achieve this end wider use must be made of the concepts underlying the quality control techniques. The frequencies of work duties or elements must always be described in terms of standards and the range of permissible variation, with the latter corresponding to the usual control limits in the quality control charts. The actual standard frequency must correspond either to the outer limit or should be close to it as it represents the maximum demand made on the worker by the particular work duty. Where frequencies exceed these limits, obviously substandard conditions must be declared to prevail, and these circumstances must be dealt with in the manner in which all substandard conditions are handled under a specific collective bargaining agreement.

The use of this approach to the problem of work assignment determination will also provide an additional incentive to management to attain not only higher levels of performance but also a narrower band of variations on a specific job. The time study technique is in practice unable to provide realistic adjustments of this type.

The TWUA Research Department has devised another approach to the problem of work assignment determination which seeks directly to apply these concepts. The benchmark or model job procedure builds upon collective bargaining concepts. Management and the bargaining agent agree upon the use of an actual job as a guide or model for the determination of all subsequent assignments on the same operation. In practice, after the parties have agreed upon such a benchmark or model job, it is described in terms of its major activities. Empirically determined conversion factors are employed to convert the frequency of each of the major activities into a single equivalent unit and the maximum permissible number of equivalent operations or activities which

a worker can be expected to perform in any stated period such as two or four hours is clearly stated. While the job assignment is set in terms of the standard frequency of equivalent operations for the test period, a maximum frequency is also set as the ceiling beyond which the variation may not proceed without making the job substandard.

The above contribution to the art of job assignment determination has the advantage of defining the job completely in terms of demands made upon the worker and therefore conforms with the management's basic desire to set jobs in terms of identical levels of worker application. It not only acknowledges the principle but actually provides an automatic procedure for insuring the enforcement of the principle since the worker under these circumstances has a personal incentive and reason to keep a check and enforce the job standard. The procedure has another advantage in that it circumvents other deficiencies of the time study procedure such as its assumption of operational stability, uniformity of motion patterns, ability to rate, the ability to project new jobs through the sum of allowances for the component parts of a job. Moreover, it provides a description of a complete job in terms of realistic physical duties and frequencies. Because these can be negotiated many of the suspicions can be moderated and opposition to measurement dispelled. Moreover, it tends to favor the mills with the most stable operating conditions where the compatability of benchmarks and quality control installations are obvious. Both incorporate a philosophy which recognizes that variation cannot be eliminated merely by ignoring it and that it must be policed and appraised by continuous sampling. Mills employing quality control would find benchmarks particularly attractive as their technicians can devote themselves to narrowing the range of variation with the knowledge that success will be reflected in increased machine assignments and productivity. Cumbersome and expensive time study analyses which even after trial installation are frequently found unsatisfactory because the effects of variation, previously overlooked, often render forecasts inaccurate can be eliminated. The high probable return per dollar invested in quality control in benchmarked mills provides compelling incentive for its adoption.

CONCLUSION

Conditions in textile mills are in constant flux, with results diverging from those expected to stem from the original raw material and machine setting specifications. Numerous variables are constantly affecting product characteristics and demands made upon the worker. Standards cannot realistically describe the job with single values. They must include in addition, a definition of the range of permissible variation. Only such definitive specifications will permit practical testing of conformance between standard and reality.

A major failing of the current time study techniques is the use of average values as absolutes. These inadequately describe the actual textile mill job demands. Jobs, therefore, are not being set in terms of actual work demands and improvements cannot be effected until quality control techniques are implemented. Unfortunately, current time study philosophy is the direct antithesis of that underlying quality control. It is therefore particularly difficult to operate both plans in conjunction with one another in textile mills. Quality control, being scientific, cannot be altered to make it more palatable to time study advocates, who, unfortunately, are more firmly entrenched in management

hierarchies. The latter are not especially prone to the acceptance of scientific techniques especially if they complicate observational and computational procedures. We have therefore offered an alternative approach in the benchmark plan under which jobs are defined in terms of equivalent operations which can be directly related to control chart recordings. Any work assignment appraising technique which relies upon evidence collected through quality control applications is more likely to produce realistic conclusions and reduce the frequency of time study bred work assignment grievances. A mutually agreed benchmark system of work assignment is the proper base for developing more satisfactory assignments in an era of collective bargaining.

* * * * * * * * *

FOOTNOTES

(1) Solomon Barkin, "Trade Union Utilization of Quality Control Techniques," Proceedings of 7th New England Quality Control Conference, (November 1953).

(2) Solomon Barkin, Franklin G. Bishop, Sumner Shapiro, Textile Workers' Job Primer, (New York, TWUA-CIO, 1953), pp. 59-67.

(3) Ibid. 62, 63, 64,86, and 87.

(4) Ford Motor Co. and United Automobile Workers, CIO Arbitration Case, 12 LA 949.

(5) Lafayette Cotton Mills and Textile Workers Union of America, CIO Arbitration Case, 17 LA 772.

(6) Kent Woolen Co., Clifton Heights, Pennsylvania, and Textile Workers Union of America, (Publication Pending).

(7) Solomon Barkin, "Concepts in the Measurement of Human Application," Industrial and Labor Relations Review, Vol. 7, No.1, (October 1953), pp. 117, 118.

Solomon Barkin, "Diversity of Time Study Practice," Industrial and Labor Relations Review, (Publication Pending).

STATISTICAL QUALITY CONTROL IN THE PRODUCTION OF COMPLEX EQUIPMENT

Major General Leslie E. Simon
Asst. Chief of Ordnance for Research and Development

So nigh is grandeur to our dust,
So near is God to man,
When duty whispers low, "Thou must,"
The youth replies, "I can."
Emerson

The problem. At first glance, the application of statistical quality control to the production of complex equipment such as guided missile systems may appear discouraging, if not impossible. Certainly, there is little or no precedent of applications of similar complexity. Let us consider the Army's recently publicized guided missile NIKE, which is designed to protect our armies and our cities by attacking in robot-like fashion and high precision the highest speed, highest altitude enemy aircraft, and by destroying them, before they can drop their bombs on the target. This missile system, according to AT&T publications, is manufactured by the Western Electric Company, the Douglas Aircraft Company, and over 1,000 subcontractors in 20 states, and requires approximately 1,500,000 parts in its highly complex control system. The application of statistical quality control to NIKE is made additionally difficult in that the item must give adequate performance (including an acceptable percentage that function properly) under world-wide geographical environments, as well as its self-generated environments ranging from sea level to the thin and frigid stratosphere.

The Attack on the Problem. In dealing with highly complex equipment we choose to use the broader term, reliability, rather than the more conventional term, quality control. We choose the former term, not only because it more nearly defines our ultimate goal of economic design and production of systems that have a high probability of functioning as intended (which includes considerations of redesign), but because we wish to make it clear at the outset that we must adapt to our use any applicable techniques, regardless of the field of science or engineering in which they may be found. However, it is astonishing to find how much of this large-scale enterprise consists essentially of the tools which the versatile quality control engineer uses in his daily work.

The Army's belief that missile reliability is susceptible to scientific attack is based, in part, on experience in the application of Statistical Quality Control to Ordnance problems of increasing complexity, since World War II. An outstanding example is our quality control program on the complete round of 105-mm howitzer ammunition. Although the use of process control and sampling inspection plans developed during the early war years resulted in a satisfactory round of ammunition, the size of homogeneous lots was still unsatisfactorily small, usually of the order of 3000 - 4000 rounds. After the war, an organized attack on the problem of large uniform lots resulted in the production of lots of uniform complete rounds as large as 150,000 rounds each. The improved quality was not only accompanied by some savings in manufacture, but made possible much larger savings in the use of ammunition in the field. The major part of designing and installing the system of quality control was done under contract by the Quality Assurance Division of the Bell Telephone Laboratories. The work could hardly have been possible earlier because of the incomplete and uneven reception given S.Q.C. by industry.

The Army's optimism on the probable success of increasing guided missile reliability through S.Q.C. is further predicated on the wartime experiences of United States industry which made it susceptible to further application of statistical concepts. This susceptibility has grown, postwar, along with the general increase in engineering skill and efficiency on the part of our great American industries. The vigor and enterprise of American industry is the real basis of our optimism regarding the successful application of the same ideas to our newer Ordnance products which are of complexity never encountered before. The complexity of guided missile systems may, indeed, seem baffling; but both our youthful science of Quality Control in particular, and our youthful American industry in general, will accept the challenge for the reduction of complexity to orderly simplicity as a "must," and reply in the words of Emerson, "I can."

I do not mean to say that optimism regarding the feasibility of reliability in complex equipment gives us license to just muddle through. The solution of the problem of reliability of complex equipment depends upon an orderly, comprehensive attack. First, a complex organizational and management problem is involved. There must be an Army reliability group in the Office of the Chief of Ordnance for policy, planning, general management, and promotion. A similar group is required in the Guided Missile Center, Redstone Arsenal, for the direction of work, analyses, identification of assignable causes of unreliability, correction of deficiencies, and feed-back of information to missile system designers. There must be a reliability group at the Proving Ground, for the inspection and test of missiles as received, for the design of experiment, for planning environmental and flight tests, and for evaluation of performance. These organizational and managerial steps within the Army were taken early and are already paying dividends.

However, much more remains to be done. A very much higher degree of success in the reliability program is possible, and much more economy is possible if the prime contractors, subcontractors, and components manufacturers make reliable products in the statistical quality control sense, rather than merely screen the product by means of old-fashioned inspection acceptance techniques. In short, industry needs more and better S.Q.C. I would like to give three important and interrelated aspects of the manufacturer's quality control problem which most need prompt attention. I shall classify these as personnel, techniques, and viewpoint.

<u>Quality Control Personnel</u>. Quality Control, unlike mere inspection, is not merely a matter of techniques. An organizational and management problem is involved. It must be entrusted to capable personnel high in the organizational structure of the company. This statement, if accepted, has two implications; one of importance to the company, and the other of importance to the profession of quality control. First, with respect to the company that is deficient in quality control, it means that it cannot correct the situation merely by urging that the persons in its existing organization become quality control conscious and by plowing-in a few junior quality control engineers at the bottom. It must get itself one or more top-rate quality control engineers (or directors) and must institute a statistical quality control organization which reports directly to top management.

Second, with respect to the quality control profession, the statement means that the profession demands that an individual (if he is to go

far) have not only statistical competence but also a high degree of general technical and managerial competence.

Some persons have taken the attitude that S.Q.C. is so simple that anyone with an ordinary education can acquire the techniques and carry out successfully a complex program. This just is not so, and has been documented repeatedly. The profession must stop leaning on a few simple tools. It must become aware that it can succeed, like any other profession, only through the all around excellence of its members.

Techniques. Not only is conventional inspection an inadequate answer to the manufacturer's quality problem, but statistical quality control must be applied in the broad sense of economic control of quality by any useful technique wherever found. In doing so, the quality control engineer generally realizes that the quality of the final assembly is controllable primarily through the control of the subassemblies and individual parts that go into it. However, he may not identify so swiftly the intensity of effort that should be placed on a specific component. The answer is generally an economic one.

Let us examine this idea a bit more closely. Consider an assembly made up of n components, A, B, - - - N.

Let R_i = the reliability of the i^{th} component.

Let V = the loss that may be incurred because of failure of an assembly due to a component failure.

Then $R_T = R_1 \times R_2 - - - - - - - - R_i - - - - - - - - - \times R_n$ where R_T is the overall reliability of the assembly.

Thus, the loss due to the unreliability is $(1 - R_T)V$

Assume that through the application of S.Q.C. the original reliability R_i is improved to $R_i + \Delta_i$

Then, the new overall reliability of the assembly, R'_T, is:

$$R'_T = R_1 \times R_2 - - - - - - \times(R_i + \Delta_i) - - - - - \times R_n$$

$$= R_T + \left[R_1 \times R_2 --- R_{i-1} \times R_{i+1} - - R_n\right]\Delta_i$$

$$= R_T + R_T \frac{\Delta_i}{R_i}$$

Now the loss due to unreliability is $\left[1 - (R_T + \frac{R_T \Delta_i}{R_i})\right] V$

The gain is $(\frac{\Delta_i}{R_i}) R_T V$

This simple formula gives rise to several interesting conclusions. The only one of these four parameters upon which we can operate is Δ_i. The importance of Δ_i increases as R_i decreases; i.e., if the ratio of Δ_i to R_i is large, we should do something to achieve Δ_i. Also, the importance of Δ_i increases as $R_T V$ increases; and since R_T should be near unity, we may say merely, that if V is large, we should try to achieve Δ_i.

Thus, the swiftest "pay-off" inheres in doing something about the worst components, provided they admit of substantial improvement and are associated with the loss of usefulness of valuable assemblies. It is very important to note that one may be justified in incurring an increase in cost of several-fold in a cheap component to increase its reliability, because its failure can result in the waste of a whole and much more valuable assembly; but not nearly such a great ratio of increase in cost may be tolerated for an expensive component, because (1) the absolute cost is greater for the same expected gain, and (2) the added cost raises significantly the value of the assembly.

Before doing very much about improving R_i, one should consider such matters as redundancy; e.g., replacing the i^{th} component with two or more such items in parallel. Thus, n redundancies in the i^{th} component gives rise to a new reliability, R'_i, where,

$$R'_i = 1 - (1 - R_i)^n$$

For the simple case of replacing one component with two in parallel, we have,

$$R'_i = \left[1 - (1 - R_i)^2\right]$$

$$= 2R_i - R_i^2$$

Redundancy is not to be recommended as a general solution, especially for guided missiles; but we can demonstrate its potential for improving reliability if we consider an assembly of eleven components in series, ten of which have a reliability of 0.999 and the eleventh of which has a reliability of 0.95. Redundancy might be the best procedure if the component were small and cheap. Duplication of the eleventh component would raise the reliability of the assembly from 0.94 to 0.99, an improvement in level which might be much more expensive to achieve through quality control. The reliability formulae also lead to a somewhat general deduction. For example, we see that in some common electronic parts, it may not be economic to try to get industry to raise its overall level of quality to that necessary for expensive complex assemblies such as guided missiles. Instead, it might be better for the industry to have two classes of the part; one of the quality required for missiles and the other suitable for the radio and TV industry.

Considerations of matters such as these may not seem an integral part of quality control, but they should certainly be part of the work of a quality control engineer who is the responsible manager of a program, if he is to arrive at the best and most economical solution to the reliability problem.

<u>Viewpoint of Quality</u>. If the company that makes complex equipment institutes quality control as a management tool, a new and important viewpoint becomes possible. The conventional viewpoint of management in meeting specifications is vertical; i.e., the product is tested at each stage in the assembly and adjusted to meet requirements. At the last stage in the assembly, the product has practically acquired a pedigree comparable to that of a prize-winning pup at a dog show. One can look

up the long vertical column through which it has passed and see the degree of perfection with which it has been checked off at each station. Management takes a great pride in system such as this, that yields a final complex assembly that so obviously meets requirements. Management will consider the suggestion of instituting Statistical Quality Control almost as an insult. What is wrong with the system?

The stage inspections and tests insured only that the assembly at hand was correct. It did not assure that the flow of components at any stage were of a satisfactory level and of uniform quality with respect to themselves. In fact, the i^{th} component may have had to be adjusted far from the norm, in order to compensate for the accumulation of tolerances allowed in its predecessors.

It unfortunately happens that we have ever with us a parts replacement problem in items of materiel accepted for Army field use. Complex assemblies, such as guided missiles, have many ills to befall them between the assembly line and their final discharge on their nation-protecting mission. These ills frequently create defects that did not exist at the time of acceptance. Hence, the assemblies must be tested at various times in their lives, and defective parts replaced with spare parts. An exchange of parts, when the parts are not of controlled uniform quality, may be tantamount to a chain reaction that throws a large part of the system out of adjustment.

It is not enough to be able to look up the vertical column and see how the assembly passed inspection at each stage, as each component was added. It is necessary--perhaps even more important--to be able to look across the horizontal row of each component and see if the quality of components is uniform. This is precisely the function of Statistical Quality Control. In short, the vertical viewpoint yields an item which is analogous to the handmade article or selective assembly that was practiced by industry before the modern concept of interchangeability of of parts. We must have the parts of all Ordnance materiel interchangeable among themselves, and in complex assemblies this can be achieved effectively and economically only by Statistical Quality Control in assembly, and Statistical Quality Control in the manufacture of the components. The work of the Ordnance Corps in the manufacture of complete rounds of ammunition has shown that this can be done and that it will result both in more economical production and in improved performance.

<u>Where We Are Now</u>. The Army's organization for reliability of complex assemblies, including quality control, has been working for some time. Some of our contractors are strongly supporting the program. Both the electronics industry and the Aircraft Industries Association are giving the reliability program much attention, and the Department of Defense is lending great stimulus to the program. Acceptable reliability is a current achievement, but much remains to be done. Statistical Quality Control as a managerial tool needs to be recognized, adopted, and instituted with authority by more contractors and manufacturers of components. Leading quality control engineers need to recognize the importance of administrative and managerial competence and their obligation to face the total problem of economic quality at every stage in the life of the product with the exercise of all the ingenuity and capabilities of the quality control profession.

QUALITY CONTROL AT WORK IN AIRLINE ACCOUNTING

Winston C. Dalleck
United Air Lines, Inc.

For a good many years statistical quality control, or more properly quality control, has been employed with outstanding success in the field of manufacturing. Quality control tools have helped to bring a new objectivity and perspective to the problems of quality, inspection and cost reduction as related to production activities. These same problems are faced by accountants and auditors. Quality relates to clerical efficiency. Verification is the inspection process in accounting and auditing operations. Cost reduction represents better and cheaper ways of getting the job done. The fundamental theory of quality control applies with equal effectiveness here as it does in related manufacturing processes. The many successful applications made in the past verify this. Even so, the field of accounting and auditing is still virgin territory for quality control.

Through the medium of the American Society for Quality Control, meeting here this week, we can exchange quality control experiences. The joint discussion of our problems and applications will encourage investigation and research into new and broader uses of these statistical techniques. The inclusion at this convention of a series of papers dealing with quality control, as applied to administrative and accounting problems, will do much to extend the horizon of statistical applications. It is a privilege to participate in this series of presentations which, it is hoped, will be repeated on a larger scale in the future.

At United Air Lines quality control techniques have been employed to improve the quality of accounting and auditing processes and also to save clerical manhours. The standard tools of control charts and acceptance sampling have been used successfully to accomplish this. In addition, we have endeavored to use statistical sampling methods as a substitute for a 100% accounting operation. This last has been considered black magic by some and as a violation of accounting convention, but we have shown that it works.

Three specific applications covering a variety of problems may be of interest to you:

1. Verification of the fare computation on passenger tickets.
2. Audit of stores inventory physical count.
3. Settlement of interline accounts between carriers.

Verification Of The Fare Computation On Passenger Tickets

At first glance it might not seem that the entry of a fare figure on your plane ticket is an involved operation. The first class one-way fare from Chicago to New York is $45.10 and round trip is $85.70. That should be a fairly simple and error proof entry to make, and it is. However, though there are many tickets sold each day for Chicago to New York passage, there are also many more sold which involve complex routings and tariff restrictions. The number of routing possibilities are almost limitless and there are tariff requirements which must be referred to for all these routings. This creates a situation with error potential. In spite of this problem it is essential that the correct

fare be entered on every ticket. It is important because (1) the recorded fare is the basic statistic used in developing passenger revenue figures for the company, and (2) it directly relates to our customer relations.

In verifying and monitoring the accuracy of fare information on tickets we are concerned with two operations involving two sets of personnel. First, at the selling source, i.e., the downtown sales office or the airport terminal ticket counter, it is important that the agents are properly trained and equipped to perform this operation. Second, in the revenue accounting office it is important to have efficient pricing clerks to audit the fare figures and to perform such additional operations as pro-rates and refunds relating to segment portions of the total ticket.

At each of these sources the control chart is employed effectively as a tool for control. Daily samples of 25, 50 or 100 tickets (as required) are taken and the error level is recorded on an np chart which is maintained at each location or in each area of operation. In addition, a detailed record is made of the frequency of errors both by type and by employee. The control chart indicates how the daily error level compares with the process average and its related control limits. The detailed record of errors is useful as the basis for locating the causes of errors, particularly when this level is high or out of control. Situations involving insufficient training, misinterpretation of new tariff regulations, volume peaks and mis-assignment of personnel are detected quickly and prompt remedial action is taken.

These applications are typical of many others which our company has employed successfully to control and improve clerical processes. For example, in the audit of air freight invoices by the Revenue Accounting Department the control chart has been used effectively to reduce error levels and thereby to cut substantially the volume of correspondence relating to differences in freight charges. Another effective application is made in the field where air freight bills are prepared originally. Here again, a control chart helps to keep down all types of clerical errors which occur in writing up a freight shipment.

In these, as well as all other applications, the control chart has provided management with a device for objective measurement. Realistic evaluation of the average and variation characteristics of a clerical process can be made. Irregularities or "assignable causes" can be located quickly and accurately. The result is good management, sensible corrective action and understanding between the employee and his manager.

Audit Of Stores Inventory Physical Count

This application involves the use of an acceptance sampling plan in auditing the accuracy of an inventory counting operation. It is an interesting problem for two reasons. First, in the initial analysis conducted in developing the sampling plan it is demonstrated, through simple statistical reasoning, that percentage sampling is ineffective as a basis for test audits. Second, the application of statistical sampling (a continuous sample in this particular problem) provides adequate controls with a minimum requirement of audit manhours.

Here is the problem in brief. United Air Lines Stores Inventory in

San Francisco contains over 61,000 items in 64 major inventory classes. These classes vary in item size from 5 to 6,000, with only 10 having fewer than 100 items. A 100% physical count of all items is made once each year by requirement of the Civil Aeronautics Board. This count is made on a scheduled basis, one inventory class at a time. A perpetual inventory record is kept on IBM cards and after each class is counted a reconciliation is made with all differences being listed. This list contains all possible errors. These errors fall in two major categories, (1) paper work errors and (2) counting errors.

The Auditing Department has a responsibility for control, i.e., for monitoring the accuracy of inventory balances. This was accomplished in the past by making test counts after the inventory teams had completed their physical count in each class. The samples taken for this purpose varied in size from 2% to 4%. A check on the efficiency of counting was thus maintained and all errors which were discovered were adjusted before reconciliation was made with the perpetual inventory records. Subsequent to the test counts, the IBM list of differences was analyzed and all major differences ($\geq \pm$ \$15.00) were adjusted. Most of the items on the difference list were paper work irregularities. However, the counting errors which were not found during the test count also appeared on the difference list. Thus, it was found that the level of counting errors in some inventory classes was higher than desirable. This fact had not been discovered by the audit test count, particularly in the smaller inventory classes.

This situation did not provide desired audit control and a better test method was sought. A statistical sample was suggested as a basis for improving the audit test and an analysis was undertaken to develop the best sampling plan. During the initial study of the problem it was demonstrated rather quickly why the previous test counts were not entirely reliable. The percentage samples from small populations result in small aggregate sample sizes with the result that sampling reliability falls off rapidly as the number of items in the sample becomes very small. This is indicated in the following table and in chart I which show some representative figures for 5 typical inventory classes and the related operating characteristic curves for each sampling situation.

Typical Inventory Classes

	A	B	C	D	E
Total					
Items in Inventory	876	2718	450	1134	161
No. of Counting Errors	17	40	17	31	4
% of Counting Errors	.019	.015	.038	.027	.025
Test Count					
Items Counted	20	44	8	26	5
Percent Counted	.023	.016	.018	.023	.031
No. of Counting Errors	0	0	1	0	0
% of Counting Errors	0	0	.125	0	0

The above table shows that the test count in these classes varies from about 2% to 3% and the aggregate sample sizes vary from 5 to 44. This situation results from the variation in the size of inventory classes.

For the smaller classes a fixed percentage sampling policy obviously will produce smaller aggregate samples. Sampling error begins to increase sharply as the aggregate sample size falls below 30. In chart I the operating characteristic curve for each inventory class is shown indicating the difference in reliability between the small and large sample sizes.

A further consideration in the problem was how to best use the IBM list of differences. Since this list contained, for each inventory class, all the errors of counting and paper work, it could be used as the sampling population for test audit purposes. This would work if the following two conditions could be satisfied:

1. It would be required that no compensating errors would occur. In other words it should be assured that no physical counting errors would be offset exactly by paper work errors.

2. Further, it would be necessary that the IBM list of differences be available for analysis soon enough after the physical count to provide adequate audit control.

A careful look at past experience indicated that the chances of exact offsets were practically nil. It was further determined that the IBM list of differences could be available and analyzed within 72 hours after the original physical count of each inventory class.

Based on the analysis of all the factors in the problem, a sequential sampling method was suggested as the best plan to satisfy the auditors' control requirements and which would be adaptable to their flexible work schedule. The IBM list of differences was to be used as the sampling population and random samples would be drawn therefrom. The following outline was suggested for analyzing and adjusting the inventory and the required sampling plans were developed.

A. Check all items of $100 or more.
B. Sample all items $15.00 to $99.99 using Plan I. (See chart II)
C. Sample all items less than $15.00 using Plan II. (See chart III)

The basic requirements of this audit program were as follows:

1. Adjust all major differences per the IBM difference list.

2. Determine, with minimum audit manhours, if the accuracy of the physical count is good or poor, i.e., acceptable or not acceptable.

These sampling plans have satisfied this requirement. Audit manhours have been reduced by as much as 50% and satisfactory control has been maintained. In fact, by being able to specify the amount of precision and the risk levels desired, we have greater confidence in our results than before. As an extra dividend this application has demonstrated the effectiveness of statistical sampling and an enlightened look at testing and sampling operations has resulted.

<u>Settlement Of Interline Accounts Between Carriers</u>

In discussing the previous two applications, attention has been directed toward the standard or popular quality control tools, i.e., the

control chart and acceptance sampling. These are the scientific sampling devices which can be applied to verification or inspection problems of the accountant and auditor. Verification is as important in accounting and auditing operations as inspection is in manufacturing. It is here that quality control applications are normally found.

There is another area in the accounting field, however, which has important potential for quality control methods and should not be overlooked. The reference here is to the type of accounting operations where a sampling method can be substituted for an actual operation. An example of this would be the detailed counting of large volume and low value inventory items where the cost of getting an exact answer would be considerably more than any reasonable and expected error. In general, this relates to problems encountered in manufacturing where 100% testing or inspection is prohibitive or uneconomical. In such cases a 100% operation is not practical. Similar situations arise in accounting activities. Actually, there probably are no situations in accounting where 100% operation is strictly prohibitive (in the sense of being destructive) but there are many which are uneconomical. The airlines have at least one such problem. It is the problem of settling interline accounts between carriers.

Briefly, this is the situation. Interline settlements between carriers arise where a passenger ticket is sold by one carrier (who also collects the money) and the transportation service is performed by some other carrier. The performing carrier must attach a price to the ticket picked up from the passenger (a "lift" in accounting parlance) and bill the selling carrier accordingly. These "lifts" of interline tickets are usually generated when a passenger uses more than one carrier in his travel itinerary. Since only the total fare is shown on the ticket, irrespective of the number of trip segments traveled or different carriers used, the fare value for each portion of the ticket is not available. The carrier performing the transportation must, therefore, price the coupon before he can bill the selling carrier. The pricing of these tickets is complicated by pro-rate requirements and tariff instructions. The result, of course, is a time consuming accounting operation.

These billings from one carrier to another are referred to as the "receivables" for the billing carrier and as "payables" for the billed carrier. In addition to the above pricing operation, each carrier performs an audit of its "payables" to verify the pricing results of the billing carrier.

This interline accounting operation poses a major problem for the accounting people of the airline industry. Passenger travel by the scheduled air carriers is growing substantially each year. In 1954 the industry will carry approximately 32,000,000 passengers, an increase of about 9% over 1953. The result, of course, is more tickets to be issued and processed. Naturally, it can be expected that the volume of interline tickets will increase proportionately which means that more people, desks and machines are going to be required each year to price these tickets and to settle accounts between carriers. For example, United Air Lines carried about 4,000,000 passengers in 1953 and expects to carry about 4,400,000 passengers in 1954, an increase of 400,000 or about 10%. This volume of passengers represents approximately 5,500,000 tickets in 1953 and 6,000,000 in 1954. Of this volume the interline

portion represents 1,600,000 in 1953 and 1,800,000 in 1954.

The magnitude of paper work handling and accounting mechanics is enormous when it is considered that each of these tickets must be processed through several operations. As the volume of air travel continues to grow the carriers recognize that the pile of tickets will grow also and, accordingly, an expansion of personnel, equipment and space will be required in ever increasing quantities.

This problem is not unique, of course, but none-the-less it has concerned airline accounting people for sometime. More than a year ago the revenue accounting people of the scheduled air carriers began discussing the possibilities of scientific sampling as an approach to the interline problem. The successful application of quality control by United and others provided an encouraging lead. Through the Revenue Accounting Committee of the Air Transport Association (a voluntary association of scheduled air carriers for mutual benefit) a sub-committee was constituted to study the potentialities of statistical sampling. This sub-committee is comprised of the following airlines: Capital, Frontier, Northwest, Trans World and United. Extensive tests and analyses have been made, with United directing the research. The results of the studies have been very satisfactory and demonstrate that a random sampling pricing method can be employed to estimate the amount of interline billings to be settled between carriers.

The approach to the problem is a random sampling plan with the samples to be drawn from the ticket population. This population is the volume of tickets which represents one month's interline business for one carrier. An example of the sampling population would be the total tickets for March, 1954 picked up by United Air Lines and which had been sold by Northwest. From this sample the following two alternatives are available as a basis for estimating the population:

1. The average rate per mile for all tickets in the sample = total fare ÷ total miles. Assuming certain specifications for precision and reliability this average rate per mile from the sample is the estimate for the total population of tickets. This estimated rate per mile is used as a multiplier (with total miles from the ticket population) to estimate the total fare value of all the tickets. The total miles are readily available through an IBM gang-punching device. Here is an example from a pilot study:

United's Interline Billings To TWA
November, 1953

	Actual	Sample	Estimate	Est. ± Act. Amt.	Est. ± Act. %
Tickets	8,196	404	-	-	-
Miles	6,032,610	296,059	6,032,610	-	-
Fare	\$316,669	\$15,520	\$316,229	-\$440	-0.1%
Avg. RPM*	\$.05249	\$.05242	\$.05242	-\$.00007	-0.1%

Est. Tot. Fare = miles x RPM = 6,032,610 x \$.15242 = \$316,229

*RPM - Rate per Mile

2. The average fare per ticket = total fare ÷ total tickets. Here the average fare per ticket becomes the estimate for the total population. Multiplying this by the total tickets in the population will give an estimate of total fare value. An example follows:

United's Interline Billings To Northwest
November, 1953

	Actual	Sample	Estimate	Est. ± Act. Amt.	%
Tickets	4,332	228	-	-	-
Fare	$167,241	$8,888	$168,852	+$1,611	+1.0
Avg. FPT*	$38.61	$38.98	$38.98	+ $0.37	+1.0

Est. Tot. Fare = tickets x FPC = 4,332 x $38.98 = $168,852

*FPT = Fare per Ticket

There are certain characteristics relating to route patterns and types of service which are inherent in the distributions of each of the above rates. These characteristics provide a stratification outline which can be used to improve sampling reliability. After studying both the RPM and FPT approaches it was found that the rate per mile method was most feasible. This method provided the best solution to the mechanical problems of stratification and sample selection.

The basic restriction imposed by the carriers on the acceptability of a sampling plan was that the maximum allowable error would not be greater than the cost of 100% pricing. Tests made demonstrated that this requirement could be satisfied and that on a cumulative basis for one year the expected difference would be less than 5/100%.

Since the average RPM, computed from each sample, is an unbiased estimate of the true RPM value in the population, it can be expected that cumulative differences between estimate and actual will approach zero. Thus, the monthly variations between estimate and actual which do occur will tend toward offsetting each other. It is reasonable to assume therefore that on a cumulative basis, say one year, the errors or differences due to sampling will be nil. As a result the reduced clerical requirements for interline ticket pricing, using the sampling method, can be considered a full reduction of operating expense.

For the major carriers participating in this sampling plan the operating expense savings could be as high as 90% of the total present cost of interline ticket pricing. For the smaller carriers this savings would range as low as 50% because their smaller ticket volumes would require larger percentage sample sizes. In addition, there are intangible benefits which will result. The interline audit will be reduced substantially. Correspondence between carriers to settle re-billings and exchanges will drop considerably.

The final step required in implementing the sampling method will be a simultest to be conducted for a period of 4 - 6 months. During this period a number of carriers will participate in performing the sampling operation in conjunction with the present 100% pricing operation. This simultest will begin about June 1, 1954. Two purposes will be served

primarily in this final test:

1. The basic assumptions for this sampling problem will be verified by a larger and broader population. This will enable the refinement of sampling specifications where necessary.

2. The procedural steps necessary to operate the sampling plan will be worked out in detail to insure standard and uniform practices for all carriers. This relates particularly to problems of sample size, random sample selection and the computation of estimated billings between carriers.

At the conclusion of the simultest period, about January 1, 1955, it is expected that the sampling method will be adopted. It is not likely that all scheduled air carriers will participate in this plan at the beginning. Full participation is not essential to the operation of the proposed sampling method. Any pair of carriers can adopt the sampling basis for settling their interline accounts. However, maximum economy will be achieved only if all of the carriers agree to participate. It is expected that this will be attained ultimately.

The idea of settling money accounts on a sampling basis is somewhat revolutionary in accounting fields. Reliance on the effectiveness of scientific sampling in this area is generally viewed with skepticism. This skepticism is a natural and healthy point of view. Applications of quality control to problems of this type must be statistically sound and, at the same time, consistent with good accounting practices. Nevertheless, the problems which face the accountant and auditor, as the mountain of paper work continues to grow, should encourage the use of the new and better tools which are available. It is hoped that the investigation made by the scheduled air carriers into the potentialities of sampling for interline settlement purposes is a forward step in this direction.

CHART I

Operating Characteristic Curves

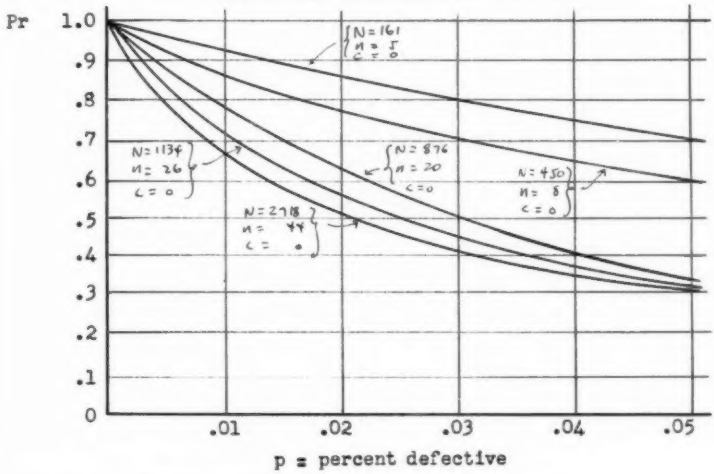
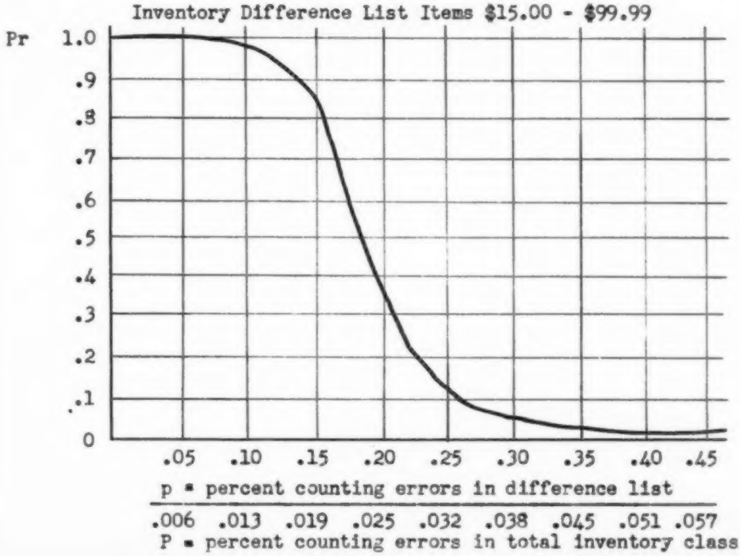


CHART II
SAMPLING PLAN I - OC CURVE

Inventory Difference List Items $15.00 - $99.99



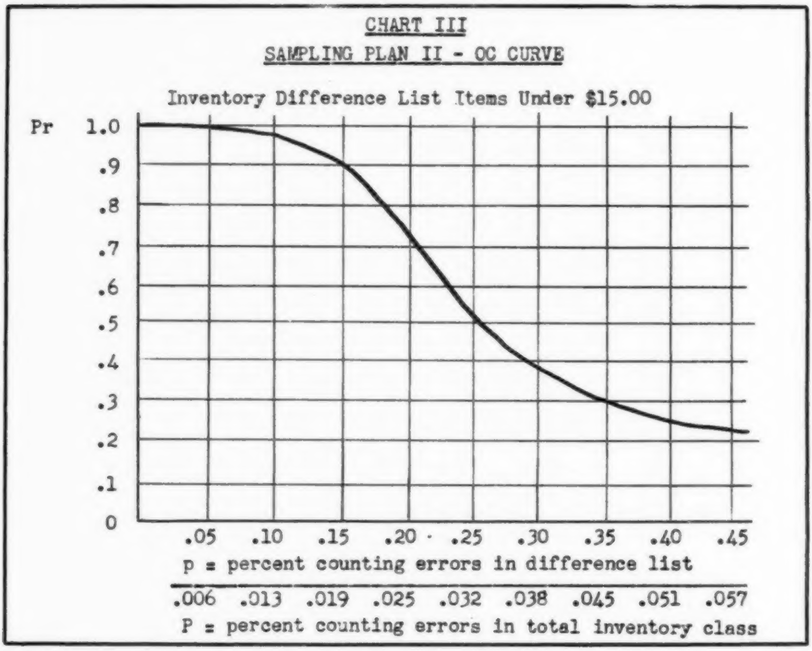


CHART III
SAMPLING PLAN II - OC CURVE
Inventory Difference List Items Under $15.00
Pr
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
.05 .10 .15 .20 .25 .30 .35 .40 .45
p = percent counting errors in difference list
.006 .013 .019 .025 .032 .038 .045 .051 .057
P = percent counting errors in total inventory class

THE AIR FORCE PROCUREMENT QUALITY CONTROL PLAN

Colonel Carl F. Damberg
Headquarters, Air Materiel Command
Wright-Patterson Air Force Base, Ohio

GENERAL

Some months ago, when I was asked to talk at this meeting on the Air Force Procurement Quality Control Plan, I accepted at once. I welcomed the opportunity to tell you about the Plan. A major reason is that I consider the adoption and use of this Procurement Quality Control Plan a most significant development in the steady forward progress of Air Force Quality Control. We in the Air Force, feel it enables us to do a far better job. We believe the Plan helps you in industry as well. But let us get into the Plan and you can judge for yourselves.

Since the Plan is an integral and vital part of the overall Air Force Quality Control Program, I propose first to discuss briefly the overall Air Force Program. Then, I shall show how the Plan fits into our system of Procurement Quality Control. Following this, I will describe the Plan and how it works. Then, I intend to draw some general conclusions as to the Plan and its impact upon both the Air Force and industry.

THE AIR FORCE QUALITY CONTROL PROGRAM

The Air Force Quality Control Program encompasses the maintenance and supply of Air Force materiel as well as procurement. The objectives of this program are to obtain and to maintain the quality of Air Force materiel needed for the Air Force to perform its mission. In this day and age of nuclear weapons and air power, the importance of attaining these objectives need hardly be emphasized. Procurement quality control must assure conformance to established quality standards in the aircraft and related equipment bought by the Air Force. Maintenance quality control must insure that quality of materiel is maintained, whether in service or when overhauled, repaired or modified. Supply quality control must look to maintaining the quality of materiel in supply channels, whether being received, stored or issued. In essence, the Air Force Quality Control Program covers Air Force materiel from the "cradle to the grave," from the day the design of an aircraft or a piece of equipment starts to take form to the day its need or usefulness has ended.

A well-known precept in quality control is that you cannot inspect quality into a product but must build quality into it. By the same token, we can say you cannot maintain quality in a product, or an item of equipment, unless it is there when the article is procured. Quality must be obtained when the item is being built. This points up the reason why we attach so much importance to the Air Force Procurement Quality Control Plan. This shows why Air Force thinking directed towards better quality led to developing a standard Plan. It is clear, Procurement Quality Control is essential to the Air Force Quality Control Program embracing, as it does, the fields of procurement, maintenance and supply.

At this point, a word of clarification may be desirable relative to what I have called the Air Force Quality Control Program. Overall management responsibility for all materiel and services is assigned by Air Force regulation to the Air Materiel Command. This responsibility includes the procurement, maintenance and supply of Air Force materiel. Responsibility for the quality control aspects of the AMC mission are assigned to the Office of Quality Control, Air Materiel Command. This office establishes quality control policies and procedures and directs, in effect, the Air Force Quality Control Program. I felt this explanation might assist you in understanding the relationship of the Plan and of Procurement Quality Control to the Air Force Program.

AIR FORCE PROCUREMENT QUALITY CONTROL

Our Procurement Quality Control system is based upon inclusion of our general specification for quality control, MIL-Q-5923B, in most procurement contracts, excluding contract studies and surveys. This specification outlines the general requirements and provides a uniform standard for an acceptable quality control system on the part of an Air Force contractor. Having made an acceptable quality control system a contractual requirement, it is the function of Air Force Procurement Quality Control to insure compliance with the contract. To accomplish this, the Air Force conducts surveillance over the contractor's quality control system (1) to determine the effectiveness of his system and (2) to ascertain that supplies and products accepted by the Air Force do conform to contractual requirements. Air Force Quality Control depends upon the contractor to obtain and record objective evidence of quality which is subsequently evaluated and verified by our Air Force personnel.

The surveillance I referred to may be considered the plan or procedure used by the Air Force to determine the contractor's compliance with 5923B and the conformance of his product to specification. When a contractor's quality control system does not fully comply with 5923B, it is evident the Air Force must do more product inspection to verify evidence of quality until compliance is effected. The nature and extent of Air Force surveillance will vary dependent upon the existence of objective evidence of quality or the lack of it. What looks like a good quality control system may be ineffective if the contractor fails to follow his own planned procedures.

The need for a complete, yet flexible, plan to be used by the Air Force in conjunction with surveillance inspection and the administration of 5923B became very apparent. A general guide towards preparation of such a plan was issued in February of 1952. While this guide did provide in broad outline form the procedures and records necessary for a complete plan, the information provided by the resulting plans did not readily lend itself to management analysis. Consequently, a decision was made to go to a standard Procurement Quality Control Plan.

THE PROCUREMENT QUALITY CONTROL PLAN

Air Materiel Command Manual 74-21, entitled "Procurement Quality Control Plan," was published last June and its implementation is now required of Air Force Quality Control personnel located in plants and facilities throughout the nation. It has attained wide usage and is enabling us to do a better job than was done in the past. We feel it benefits industry as well as the Air Force. I am sure you will

agree when I have finished telling you how the Plan operates.

The Plan provides efficient, orderly procedures and standard record forms for use by Air Force Quality Control personnel responsible for surveillance over a contractor's system of quality control. Used properly, 74-21 covers every important element of the contractor's quality system. Areas where the contractor is not acquiring objective evidence of quality can be determined readily. Failure of the contractor to act upon evidence of recurring deficiencies is brought to light. Generally speaking, careful and thorough use of the Plan gives the Air Force the information it needs to determine whether or not the contractor, in fact, has an effective and economical system of quality control and whether his product meets the required quality standards. The Procurement Quality Control Plan is broken down into five sections. These are:

Product Identification
Allocation of Inspection Manpower
Air Force Product Inspection Surveillance
Special Systems Surveillance, and
Analysis and Action

In the appendix (Tabs A, B, and C) of this paper are copies of some of the forms used to implement these sections. These are provided to give you a better idea as to how the Plan works.

PRODUCT IDENTIFICATION

The first section of the Plan, "Product Identification" provides for a systematic and thorough appraisal of the extent and type of Air Force surveillance effort required in a facility. In other words, by surveying the products and processes involved in a contractor's operations, we can arrive at a pretty sound estimate of the Air Force quality control workload. A Product Identification Chart, AMC Form 64, (Tab A) is used to record the information required to estimate workload. Listed on this form are the principal items such as complete units of simple products or parts; sub-assemblies and components of more complex end items. These principal items are treated in three general categories, namely, items purchased, items manufactured and items which have been Government source inspected and accepted.

Information about items purchased includes applicable specifications, whether bought on direct order or sub-order, whether an end item or a component, the applicability of MIL-Q-5923B, and whether or not source inspection by the Air Force is required. Vendors special processes, such as magnaflux, cadmium plating and welding applying to the item must be recorded. Items manufactured in the facility must be classified as components or assemblies, the special processing they require must be listed and an entry made to indicate if they require operational testing. Items of Government Furnished Aircraft Equipment are listed on Form 64, but only require identification as such.

After identification of the principal items which go into the final product, an estimate can be made as to the Air Force surveillance effort required. For example, we know that special processes, such as X-Ray and anodizing must be checked and that additional receiving inspection and laboratory testing surveillance are required for special processes

of vendors whose product is not source inspected. If a Government specification applies to an item, Air Force review of contractor's engineering data, testing, processing and inspection methods must be made to insure compliance. Direct order items normally will require packaging, packing and shipping surveillance.

A thorough job of Product Identification discloses what needs to be known about the contractor's products and processes in order to establish complete surveillance over his quality control system. The information obtained can be used as a guide for determining the extent of direct and indirect inspection activity required, the systems which must be checked, and the inspection stations which must be established. In other words, the workload for our Air Force quality control personnel has been determined.

ALLOCATION OF INSPECTION MANPOWER

The next section in the Plan, "Allocation of Inspection Manpower," is designed to provide the most efficient distribution of Air Force personnel throughout a facility, and to assign specific duties and work schedules to these personnel. We know the number of personnel available. We know the scope of the contractor's operation and what must be covered to effect complete surveillance over his quality control system. We may have some idea of the quality of the products being turned out from past contractor's inspection records. Having this information, we are ready to match our available inspection manpower with the existing workload.

To assist our personnel, a Form 64A, entitled "Distribution of Personnel," (Tab B), is provided. Total available Air Force manhours per month are calculated and entered on the form. Based upon the previous survey of the contractor's operations, we list as "direct" inspection or "indirect" inspection all the activity we must cover. First we determine the minimum number of manhours we must devote to "indirect" inspection activities, such as administration, review of Air Force inspection records, training, review of engineering changes and special reports. Then we must determine how much time we should devote to the inspection of systems, such as material review, tool and gage, and packaging. We allocate systems manhours first. All remaining available manhours are assigned to product inspection. While this seems a rather backhanded way to decide on the effort to be placed upon product inspection, we have found that it is effective.

Since product inspection is the major part of our direct inspection activity, I will discuss it in some detail. Initially, we determine the areas where the contractor is performing inspection and the percentage of his total inspection manhours he is devoting to each manufacturing area. By making these Air Force product inspection stations, such as receiving, machined parts, assembly and final inspection, and determining the percentage of contractor inspection effort allocated to each station, we can estimate, roughly, the percentage of the available product inspection that should be concentrated at each station. Naturally, the Air Force product inspection effort allocated will be changed when quality difficulties warrant or other considerations require it.

Guided by the contractors' inspection assignments and after consideration of other influencing factors, we have settled upon a distribution of personnel to meet our workload. Now we are ready to make out

the "Duty Assignment Chart," another form supplied with the Plan. This chart is used to assign the systems or inspection stations for which each Air Force quality control man will be responsible. The frequency of inspection and the number of hours per inspection scheduled for each system or inspection station are designated. The appropriate Air Force check lists to be used are also indicated.

PRODUCT AND SYSTEMS SURVEILLANCE

Sections I and II of the Plan have provided a standard approach for planning the work and utilizing the manpower in a facility. Sections III and IV, entitled "Air Force Product Inspection Surveillance" and "Special Systems Surveillance," respectively, describe, step-by-step, how the Air Force goes about actual product inspection and verification and reviews the contractor's procedures and special systems. Without getting into too much detail, I would like to tell you how this very essential direct inspection is performed and the results recorded.

Two forms, "Product Inspection Record" and "Procedures and System Record," are provided by the Plan to record Air Force direct inspection activity. The characteristics inspected are recorded on these forms; this is true regardless of whether the characteristics be the dimensions of a physical item, such as a pump shaft, or those pertaining to an inspection procedure or special system of manufacture. Because Government specifications prescribe the procedures and systems, it was possible to provide standard weighted check lists for them in the Plan. As a result, the percentage deficient observed during each inspection of them may be calculated and compared. However, product items, such as the pump shaft, require individual inspection treatment and check lists could not be provided in the Plan. For these a calculation of defectiveness for combined inspections is desirable and the weighting of importance is a function of any classification of defects which is employed. The records for procedures and systems and for product furnish a complete and exact report of every characteristic inspected.

ANALYSIS AND ACTION

Now comes the pay-off section of the Plan, Section V, entitled "Analysis and Action." A somewhat formidable array of four forms are used for analysis and action. The titles of these forms are self-explanatory being (1) "Procedures and Systems Record Summary," (2) "Product Inspection Summary," (3) "Quality Trend Chart," and (4) "Corrective Action Record." The summary forms provide percent deficient figures which can be plotted on the quality trend charts showing percent deficient over a period of time. The substantiating data in the individual inspection records can be quickly referenced and checked. The percent deficient is completely accurate and factual; it represents the defects or demerits found divided by the number of observations made. I stress this because there has been a feeling in some quarters that to be of value product inspection must be conducted upon the basis of a published sampling table. It is true that caution must be exercised with respect to inferences and decisions regarding the similar product not inspected and, therefore, not covered by the records.

The requests made to the contractor for corrective action are effectively supported by the quality and completeness of the individual

record data. Corrective action requests, commitments, and follow-up activity are all recorded on a form provided for that purpose. As a result of all this planning we can review the quality status and trends of systems, procedures and products. The importance of careful evaluation and analysis of accumulated data is repeatedly emphasized and fully exploited. Most important is the fact that the data is immediately used to assist inspection personnel to more efficiently conduct their activity. It is no longer considered merely as a historical record to be used or ignored as one sees fit. Analysis must be made with a view to determining action to be taken.

BENEFITS OF THE PLAN

Having described the Air Force Procurement Quality Control Plan, I would like to tell you what the Air Force and industry stands to gain through Air Force use of such a standard plan. We feel that its careful implementation by Air Force personnel definitely will show how well the contractor's quality control system complies with the general requirements of MIL-Q-5923B. Further, the fact that one basic plan is used regardless of the facility, the product or the industry involved provides the Air Force with a very effective management tool. The Plan can be and is being operated within existing personnel and budgetary limitations. Experience gained from effective application in one plant is applicable to other plants manufacturing the same or a similar product. Our administrative and supervisory personnel are provided a means for comparing the effectiveness of our various surveillance activities and for judging the comparative workload associated with the different phases or categories of surveillance effort. Better utilization of our assigned personnel is attained and made more effective. Costly and needless duplication of contractor inspection is eliminated.

Industry also benefits from Air Force use of the Plan. Use of a standard plan in connection with MIL-Q-5923B means all Air Force contractors receive the same fair and equitable treatment. The requirements upon the contractor do not vary dependent upon the individual desires of Air Force personnel assigned to a plant. Objective evidence of quality is required regardless of the industry or product. Each contractor knows surveillance effort will be directed and concentrated where needed most. Those features of the Plan which utilize contractors' inspection records, inspection instructions, and statistical sampling procedures and data serve to increase the ability of industry to communicate with the Air Force in concrete, objective terms. The contractor profits by Air Force concentration of effort on recurrently unsatisfactory conditions and recognition of the relative importance of different types and extent of defectiveness. Focusing of attention on defects as they occur in process of manufacture helps the contractor reduce his rework, repair and scrap rates. Verification inspection information recorded and used by the Air Force is readily available to the contractor as additional data for checking his own quality control system.

I probably could list many more benefits, tangible and otherwise, to both the Air Force and industry. Rather than belaboring the point, I feel we should agree that many of the benefits derived from our Procurement Quality Control Plan are mutual. Air Force efforts supplement and complement those of industry. We share a mutual interest in turning out a quality product. The Plan enables industry and the Air Force to

work side by side in attaining the quality of aircraft and associated equipment, despite increasing complexity, which the Air Force must have to meet the increasing responsibilities being placed upon airpower.

SUMMARY

Now, I would like to summarize the main points of this talk. First, the Air Force Quality Control Program administered by the Air Materiel Command views quality control in broad perspective and is concerned with the quality of Air Force materiel from its development to retirement from active use. Next, Procurement Quality Control plays a vital role in the overall Air Force Quality Program. Our Procurement Quality Control system places the responsibility for providing an acceptable quality control system and ultimate quality of product squarely upon the contractor. The Air Force, through use of Air Materiel Command Manual 74-21, entitled "Procurement Quality Control Plan," effectively assures compliance with MIL-Q-5923B and the conformance of product to specification. Lastly, the Air Force and industry share mutually in the substantial benefits accruing from the Plan, the results of an ideal industry-Air Force relationship.

Let me again express my appreciation for affording me this opportunity to tell you about the progress your Air Force is making in the field of Quality Control. On behalf of the Air Materiel Command, I should like to wish the American Society for Quality Control continued success and growth.

Tab A, Form 64
Tab B, Form 64A
Tab C, Form 64D

PRODUCT IDENTIFICATION CHART										APD					APR				FACILITY			
1	2	3		4	5		6			MANUFACTURED					7		PURCHASED					
										A SPECIAL PROCESSES					A SOURCE INSP		B NON-SOURCE		C VENDORS SPECIAL PROCESS			
ITEM PURCHASED	SPEC	DIRECT ORDER	SUB ORDER	MIL-Q 5923 APPL	END ITEM	COMP PARTS	COMPONENTS	ASSY	OPNL TEST						COMPONENTS	ACCESSORIES	COMPONENTS	ACCESSORIES				

TAB A

DISTRIBUTION OF PERSONNEL

DATE

APO

APR

PERIOD

FACILITY

1. DIRECT INSPECTION ACTIVITY	PERCENT OF INSPECTION MA				
A. PRODUCT INSPECTION	CONT	AF	CONT	AF	CON

B. SYSTEMS INSPECTION	AIR FORCE MANHOURS PER		

2. INDIRECT INSPECTION ACTIVITY			
3. TOTAL AIR FORCE AVAILABLE MANHOURS PER MONTH			
4. AIR FORCE MANHOURS PER MONTH REQUIRED FOR INDIRECT ACTIVITY			
5. AIR FORCE MANHOURS PER MONTH AVAILABLE FOR SYSTEMS INSPECTION			
6. AIR FORCE MANHOURS PER MONTH AVAILABLE FOR PRODUCT INSPECTION			

TAB B

PROCEDURES OR SYSTEMS REVIEW RECORD	APD	APR
SYSTEM OR INSPECTION STATION	CHECK LIST NO.	FACILITY

1 CHARACTERISTIC OR CODE NO.	2 De-mer-its	3 DATE							
4. Total Number of Characteristics Inspected									
5. Total Number of Defects Observed									
6. Total Number of Demerits Possible									
7. Total Number of Demerits Observed									
8. Percent Deficient									
9. Inspector									

TAB C

AN ENGINEER EVALUATES STATISTICAL METHODS

Nello Coda
Erie Resistor Corporation

It is not the intention of this paper to present a learned evaluation of statistical methods, but rather to report a few instances of my discovery and use of techniques of statistical anaylsis which, while undoubtedly very familiar to the members of this society, have been a great revelation to me and of significant help in the performance of my work. Like many other engineers who graduated twelve or more years ago I did not have the opportunity to learn statistical methods of analysis in college and was forced to rely on what is generally called common sense or judgement in arriving at various conclusions. Tests of significance and perhaps more efficient design of experiments have given me increased confidence in my work.

Before proceeding any further with this discussion, I should like to describe briefly my position in our organization and point out where statistical methods can be of help to us. I am in charge of the Electrical Engineering department whose function is to design new products, modify old ones and evaluate new material and methods as they become available. Our main products are ceramic and mica capacitors, composition resistors, and assemblies employing the above and other electrical components. Because of the extremely rapid development in our industry, we are constantly working on a large number of new projects with the attendant development and evaluation of new dielectric and resistive materials, finishes, seals and manufacturing methods. This requires that we perform a great many tests and experiments and proper design and efficient evaluation can mean great savings in time and money.

Perhaps my first introduction to the great advantages of the application of statistical technique was from the examination of the records of our Quality Control department during the first year of its establishment. For a long time we had sampled incoming material and sorted our outgoing products, with records kept in not too systematic a manner and with the result that at no time did we really have a clear picture of the compliance of the materials with the specifications. As a result of this, our specifications or tolerances were in some cases much tighter than was economically justified and in others looser than necessary. Here, then, a simple, systematic method of record keeping over long periods of time was by itself sufficient to show the need for many revisions and point the way to the writing of more realistic specifications. A word of caution, of course, must be sounded not to take or permit too literal an interpretation of this data, for unless it is analyzed properly it is very easy to reach a wrong conclusion. A rather recent experience to illustrate this point came in connection with our design and construction of a machine which applies the silver electrodes to ceramic parts to form a capacitor. As many of you know, the electrode area is very important and is one of the factors that we control in order to obtain the required capacitance value. When this machine was first set up, we decided that we should determine its process capability, so we ran it for a number of days keeping detailed records of the down time and other difficulties as well as a $\bar{X}$ and R chart of the capacitance value produced. Since there was considerable interest in the performance of this machine, people from various groups in the company followed the initial runs and in no time at all we heard that the machine could maintain a spread of 5% in capacitance which was much better than we had

expected. Needless to say, the 5% was the greatest range observed from samples of five units and the machine, while quite good did not really measure up to that.

One very good application that we have for statistical anaylsis is in the evaluation of dielectric materials, specifically, the life time of the dielectric as a function of voltage stress and temperature. This has been a very difficult task due to the large number of variables in the make-up of the dielectrics, the extreme variation in life time within a sample group, and the great length of time required to perform a test. By examining large amounts of data, we are beginning to observe normality in distributions that used to be considered quite variable and in no way predictable. The life time of capacitors under accelerating conditions of high voltage and temperature shows a rather normal distribution when plotted on logarithmic paper; in other words, the log of life time in hours or days seems to follow a normal distribution for a particular combination of voltage and temperature. This, then, has enabled us to separate the effects of voltage stress and temperature and to observe the effect of each of these variables. As we gain more experience on this subject we hope to be able to obtain valid information from greatly accelerated tests and, therefore, evaluate materials more rapidly than at present.

Another application where a statistical approach has given us some insight as to the working of a failure mechanism has been the study of the moisture seal of a certain type of capacitor. One very simple and common way of evaluating the seal of capacitors is to subject them to any one of a number of moisture treatments and then measure the insulation resistance. Unfortunately, the distribution of this parameter is usually quite large, a ten to one variation being quite common even among perfectly good units, so that the analysis of the results from a seal test seemed at first rather difficult. By plotting the logarithm of the insulation resistance on normal probability paper both before and after the exposure to the moisture treatment, we observed a very significant picture in that in addition to a significant lowering of the insulation resistance level, there was also a distinct break or knee in the line indicating the introduction of a new distribution at a much lower level. The position of this break would vary at percentage values of 5% to 30%. The interpretation of this, of course, was rather simple after the observation was made; it was evident in the first place that all units dropped in resistance as a result of the moisture treatment due to the permeability of the sealing compound, but in addition a varying percentage of seals failed because of defects giving rise to the lower distribution. Our problem then resolved itself to a study of the outright seal failures and means to reduce them. Some progress has been made and we expect to have the situation under control in a short time.

Analysis of regression and correlation is also a very strong tool which can be used when other techniques do not readily give a solution. When the number of variables is large and they are such that they cannot be controlled, it is often possible to separate the effect of each of these variables by determining the influence, if any, each has on the end result. Many times the answer is negative and no further thought need be given to their control. One case that will illustrate this point arose recently in one of our production lines where a protective coating was applied automatically to a capacitor by repeated dips, each followed by air drying and with a final high temperature cure. The nature of this coating was such that we often obtained excessive coverage

on the leads of the capacitor which necessitated removal by scraping. This problem had been chronic with us for almost a year and any number of explanations had been advanced as to its cause, such as the temperature of the bath, its viscosity, the ambient temperature and humidity and probably others. Before any money was spent, however, to control these variables, we decided to collect data over some period of time during which temperature, humidity, and viscosity varied over their normal range. Absolutely no correlation was noted between the suspected agents and the coating, but the real culprit was discovered in the variable level of the material in the pot which was thought to be maintained constant and in the variation of a positioning device. The solution was quite simple and this operation is now under control.

The examples I have cited represent but a few of the applications we have made of statistical methods, either independently or with the assistance and cooperation of our Quality Control department. In every case the rewards have been great in either time or money saved and always in a feeling of greater confidence and satisfaction. Unfortunately, it has not been easy at times to gain the required background knowledge for the use of these techniques in experimental work, since books and literature available on this subject do not generally make for easy reading. I sincerely hope that more and more colleges will introduce at least a one semester course in this subject in addition to the immediate application of these techniques to the analysis of experimental data obtained in laboratories or other courses. The members of this society can and should continue with their pioneering work by writing articles in various engineering publications and occasionally taking a little time out to educate us engineers in the wonders of their profession. Colored or numbered beads are not really necessary to explain the high sensitivity of a $\bar{X}$ chart; the simple formula relating the distribution of the averages to that of the universe should be quite sufficient. In closing I should like to recall the story of the husband who, when asked by his wife what the preacher thought about sin, answered: "he was against it", but summarize my opinion of statistical methods with the positive answer "I am for it".

A QUICK METHOD OF DETERMINING THE CHARACTERISTICS OF A FREQUENCY DISTRIBUTION

Arthur Bender, Jr.
Delco-Remy Division,
General Motors Corp.

The practical use of the Frequency Distribution Curve, as a means of illustrating variation patterns, is increasing in all types of industrial work. Usually a theoretical curve is fitted to the actual data and then a comparison is made to specifications. Considerable time is required to:

1. Make calculations for the average;
2. Make calculations for the spread;
3. Calculate and draw the smooth theoretical curve;
4. Calculate the percentage outside of specifications (using tables).

A method is presented which:

1. Determines the average;
2. Determines the spread;
3. Determines whether the distribution is normal or not;
4. Determines the percentage outside of any selected tolerance;
5. Provides a graphical means of developing the distribution curve.

The only arithmetic operations required are addition and the calculation of percentage. If the selected sample sizes are 25, 50, 100, etc., the calculation of percentage is greatly simplified.

Copies of the paper, showing complete step-by-step procedure, will be available to those attending the session.

NOTE: Mr. Bender's complete paper may be found on pages 82-86 of the "8th Midwest Quality Control Conference Papers, 1953". Copies may also be had by writing direct to Mr. Bender at Delco-Remy Division, General Motors Corp., Anderson, Indiana.

THE ROLE OF STATISTICAL QUALITY CONTROL IN TODAY'S CHEMICAL COMPANY

Maynard S. Renner
Dewey and Almy Chemical Company

"Do you use statistical quality control methods in your testing work?" is a question I have asked in many chemical laboratories during the past several years. Suprisingly often the reply has been: "Well -- we are interested in statistical quality control, we would like to be using it, but right now we aren't doing anything about it." Often the speaker adds, "one (or two or three) of our men took a night course in statistical quality control one (or two or three) years ago" or "we'd like to do something about it but somehow we never seem to get started."

This situation has always puzzled me. Arousing interest and creating desire is said to be the biggest part of selling. But here, interest has been aroused, desire created, but no sale made. Why? What is stopping it?

This paper deals with this problem. It discusses some of the reasons why people have trouble getting started using statistics; indicates a way to make a beginning; discusses some of the advantages of making a beginning; and discusses some relevant problems of specification and testing precision in the chemical industry. It is written, not by a statistician, but by a staff man who has been there himself.

Let's start by asking, "why use statistical methods? Is there any good reason?" In today's economic climate the only concrete reason for a chemical company's using statistical methods should be to reduce overall operating costs - overall costs, not solely inspection costs or spoilage costs. Evaluation should, of course, not be short term, but rather in terms of the long view that chemical companies must use to stay in business. If using statistical methods will cut operating costs they should be used. If their use will add to operating costs, they shouldn't be used.

Now how can using statistics reduce overall operating costs? Statistics is purely a tool - a tool for making decisions. Using statistics can reduce overall operating costs only as it helps an organization to make decisions more economically, that is with less expenditure of manpower or equipment in the decision-making process; or to make decisions more precisely, and thus obtain a more efficient use of raw materials. But are statistical methods not being used because they don't do these things? I think not. Certainly some well-run chemical companies, with recognizedly astute management, are spending heavily on statistical programs. Rather I think the difficulty comes from two other sources.

In the first place, chemistry itself, in industry, is a decision-making tool. As a simple example, we don't run an acid number on a sample to learn its acid number but to decide whether to accept or reject the lot it represents. Chemistry is not only a science-based tool, but is a very workable tool. Most of the tests used in our laboratories were developed without help of statistics. Our company was built upon quality and has maintained an excellent reputation for high quality products by chemical methods alone, without using statistics until recently. In other words, it may help, but you don't need statistical methods to run a chemical company.

Furthermore, the use of chemistry alone as a decision-making tool often involves a way of thinking which is very different from the way of thinking of statistics. A chemist runs a test and gets a number. This number is often considered to be, or to stand for, a "true", or absolute value. The importance of precision or lack of precision may or may not be seen, but the number obtained from the test nevertheless has definite standing as a number - maybe not a very good number but still the best obtainable with the time and equipment at hand. Often the real worth of the number never gets tested because it was obtained by the best method anyone could think of: therefore it is assumed to be the best number anyone could get under the circumstances.

Now this way of thinking, statistically unrealistic, is also of course completely incompatible with the ideas of statistics. It may be abhorrent to the statistician, but remember it can make statistics equally abhorrent to the chemist. Why should he fool around with a lot of complicated mathematics, maybe even repeat a test a couple of times - he already has his number!

This pattern of thinking is a part of the custom of many chemical laboratories, and the power of custom has been described by Ruth Benedict (1) as follows: "No man ever looks at the world with pristine eyes. He sees it edited by a definite set of customs and institutions and ways of thinking. Even in his philosophical probing he cannot go behind these stereotypes, his very concepts of the true and false will still have reference to his particular traditional customs".

Some recent developments in the field of sociology, psychology and anthropology also help to explain why this pattern of thinking should have developed such force. It is only by trying ideas out in action that a person can make the judgments that will let him find out if a given belief is an effective operating tool. Thus, we are told, (2) all of us build up an assumptive world of beliefs that we have tested in action and found to be workable. Now what happens to the chemist's belief that the value he finds is the one true value, when he tries it out? I think we all know the answer: it works - to the chemist's satisfaction - most of the time. Its accompanying by-products such as excessive re-checking, feelings of guilt when rejecting "conflicting" data, and the occasional faulty decision are accepted as necessary evils.

The answer here, in my opinion, is to communicate with the chemist, not in the language of statistics but rather in the chemist's own language - his data - to get across by continued repetition, illustration and explanation that these items of data are not magic numbers, but rather samples out of one or more populations. The chemist does recognize that there is uncertainty in his results. He is interested and often grateful for the point of view that shows his items of data neither as discrepancies, nor as standard deviations, but rather as part of a continuous and often predictable whole.

This approach speaks a language the chemist can understand. It also develops a growing appreciation for, and understanding of, the idea that variation is to be expected in nature, thus it may lead to an understanding of the concept of variability. An understanding of this concept is necessary not only for an appreciation of the need for statistical methods, but also for an appreciation of how and where to use statistical methods; where and how not to use them also. (Understanding this

concept is a great safeguard against misuse of statistical procedures; of getting the cart before the horse; the statistical technique before the problem.)

The second difficulty is the way of introducing statistical methods into a chemical company is that the average man is not accustomed to think in the language of mathematics, unless the kind of work he is doing directly calls for mathematics. Again, it is a matter of customs and beliefs that have worked. Industrial problems must be solved in a hurry, usually can be solved by thinking a language of things, materials, and equipment. The result is that the average chemist who knows only too well how to take an average, reacts with complete bewilderment to as simple an expression as the description of an average in terms of a mathematical formula as follows:

$$\bar{X} = \frac{\sum_{1}^{N} X_i}{N}$$

One solution for this difficulty, possibly not the best one, but at least a workable one, is to use very simple statistical methods. We have had good results for example with a control chart using control limits calculated from mean successive differences or "running ranges" as developed for the chemical industry by C. A. Bicking (3,4,5). These charts, requiring only simple calculations, have proven most valuable, particularly for many of our uses where we have relatively few batches of many different products. The chemist first uses them to tell him how well specifications agree with what the process is actually doing, with sometimes surprising results. Later he may learn that the calculated control limits, if the process is reasonably stable, give him a workable set of numbers to use to describe the behavior of the process.

It is such experiences that provide a connecting link between abstractly knowing about variability and the actual testing by action which is needed to convert abstract knowledge into firm belief. Gradually there develops a point of view which is a more effective operating tool because it agrees more closely with reality than the point of view which tends to look upon each discrete value by itself. It also provides a way to use all the data - a prime rule in problem solving (6). No one would think of throwing out a piece of process equipment after making only one batch in it, yet in handling testing data, all too often, without control chart or other statistical tool, all past experience is lost slight of and only the immediate moment's results considered.

Thus it might be said that the role of Statistical Quality Control in today's chemical company is a conservative role: to provide a means for examining and describing the discrete items of data of the immediate present in the light of the sum total of past and predictable future experience of the same kind of data.

Here the value is not so much in statistics as in what statistics does. Its greatest value may possibly be semantic: it provides a way to describe quality situations more completely and in objective rather than in emotional terms. It probably takes on some of the aspects and values of mathematizing as described by Korzybski (7) as follows: "Mathematizing represents a very simple and easy human activity because it deals with fictitious entities and we proceed by remembering. The structure

of mathematics because of this over-simplicity, yet structural similarity with the external world, makes it possible for man to build verbal systems of remarkable validity."

Consider, for example, what happens when specifications are examined in the light of the statistical quality control concept that the specification is economically determined on a factual basis. What quality does the customer need? What quality can the supplier make? If a compromise is necessary, which compromise will give customer and supplier the most favorable costs? How much variation must be expected in the product? In answering these questions, a slow, often subtle, but great change in an organization's habitual pattern of thinking about quality occurs. Quality ceases to be a matter of unanswerable questions of opinion. Quality questions are stated more precisely and precisely stated questions have the one great value that their answers are clearly indicated (8). Quality becomes recognized as a basic economic problem in which the customer's use requirements, the vendor's process capability and the variability of the process are the important controlling factors.

The first signs of a pay-off come in the form of tangible savings in material costs from more economic use of raw materials because of better quality decisions.

The simplest example is the oldest and applies to all products bought on a weight basis and sold on a volume basis - including packaged products sold on a fixed weight per package basis. Without control charts weight per unit (square yard or package) usually fluctuates more widely than with control charts. To maintain the required minimum level it is necessary to run at a high average. The use of control charts calls attention to the importance of weight variation and eventually reduces it so that the product can be made to the same minimum level but at a lower average value with consequent saving in material cost. Also, of course, a more constant product is more desirable to the customer than a widely fluctuating product.

This kind of saving often requires long study or a change of equipment to obtain - is too often overlooked. I think it is entirely true to say that anyone who is selling a packaged product on a guaranteed weight per package basis in substantial amount and not using statistical quality control is pouring money down the drain.

Other gains also come with control charts. People begin to watch the charts and to ask what causes some of the quality fluctuations they see. Spoilage and rework are reduced. When content or proportion of product or active ingredient is the fluctuating quality characteristic, the attempt to explain and control it can, almost without effort, lead to increased yields and hence more cost reduction. We are today using control charts in only a limited way but already our savings have amounted to many thousands of dollars per year from this source, with indications of much more to come.

But material costs are not the biggest part of a company's expenses. One accountant estimates that in a typical chemical company, material costs may vary from 28% of total expenses (9), for an unprofitable volume of business to 35% of total expenses for a normally profitable volume. Savings through the use of statistical quality control can be obtained in the other areas of expense also. One tangible and highly pub-

licized source of such saving comes through reduced sampling and testing or inspection costs. But there are great intangible savings also, large-ly because quality decisions can now be made more efficiently and more ecomomically; and making decisions in today's growning companies, with their growing communication problems can cost money.

People on decision-making levels in a company - management, sales, research, engineering, quality control - don't punch job tickets. So you can't tell how many hours spent arguing over whose opinion is right are saved by letting the data do the talking: nor how many hours of lost effectiveness due to resentments, aggressions and hostilities arising from internal friction over conflicting opinions are saved by having the data in shape so that the right action is clearly indicated: nor how much increased organization effectiveness comes from the improved habits of thought that develop through making decisions from facts instead of opinions.

One of our products that was considered very troublesome a few years ago today is considered one of our most satisfactory. Simply plotting a control chart showed the process to be remarkably uniform, but definite-ly out of step with specification limits with respect to a major though non-critical quality characteristic. Study of the customer requirements showed that it would actually be desirable to change the specifications to agree with the process capability of the product. The change was made with a resultant marked improvement in our opinion of the product, and a saving of much time spent previously deciding what to do about "off-specification" batches.

One step or series of steps in the manufacture of many of our pro-ducts, as with many chemical companies, is the process of "adjusting" to final characteristics. These adjustments of course, cost money, in time, manpower, and equipment. One of our laboratory heads in charge of this kind of work has succeeded in reducing the former average of 1.5 adjust-ments per batch to one adjustment per 10 batches largely as the result of the better perspective and improved judgments made available to him by the historical background of the control chart.

Specification meetings for products under control charts run quietly and amicably. Each process is within specifications or it is not. If it is not, then the choice must be made between changing the process to fit the specification or changing the specification to fit the process. One small group of products is recognized as varying more than is considered desirable. However, the variation is also recognized as being inherent in this particular process. So long as succeeding batches stay within the indicated process limits, nobody is going to be badly upset about it. In the meantime, the energy that would otherwise be diverted into ques-tioning and deciding about each deviating batch as it occurred, can in-stead be given to completing the field testing of a new series of prod-ucts which have less variability.

But it is important to remember that the approach described above is along plausible or what the mathematician calls heuristic lines of rea-soning instead of along the airtight lines of reasoning of the concept of statistical quality control according to Shewhart. In many ways - in many places - in its many phases, the chemical industry is different from the discrete-piece industries and it would seem that often the plausible or heuristic approach is the only way to make a beginning.

For example many chemical products made in batches are not made very often, at least as compared with the many units of production of the discrete - piece industries. For such products it may be impossible ever to get the amount of data needed to establish the high degree of control or predictability implicit in the Shewhart concept of quality control. Thus calculated control limits for many chemical processes although accepted for working purposes as Shewhart limits may be, in a strict sense, what Shewhart calls "Students Ranges" (10).

Chemical specifications may put less emphasis upon specific performance requirements than do discrete-piece industry specifications. Different customers', often the same customer's, use-requirements of material, equipment and process may vary widely. Sometimes no performance test other than actual use in production has been developed. Again use-requirements may not yet have been thoroughly established. In such cases the specifications may be written largely to give assurance by means of a number of chemical or physical tests that each batch will be the same as all previous batches.

Simons' comments regarding this general phase of the problem of assuring standard quality are worthy of note here (11): "It is impossible to specify completely a definite series of operations which will certainly detect that a very simple product is not standard quality. Some of the quality characteristics are not capable of definite measurement and after all, there is a limit to detail. Hence, the specifications of the intended quality must be limited to some chosen or principal quality characteristics of the product".

At present most of our specifications are set on the basis of engineering judgment considering all available information about process capability and customer performance requirements. Often today specific data about the customer's needs is incomplete, hard to get. Sometimes the tolerance limits required for satisfactory performance have not yet been precisely determined; other times the customer's process is not stable enough to justify setting tolerance limits - what is wanted is the vague characteristic often called "latitude".

Many of these problems will in time, be worked out. As a step in this direction, we are doing more and more field sampling not only to learn more about the quality of our product in the field but also to get still more specific data about the relationship between performance in the customer's plant and testing characteristics.

In thinking about specifications we must think also about testing precision since basically today precision and process variability are the two elements that determine how wide the specification limits should be. In discussing this subject, Simon suggests (12), "a general rule, which of course, is not universally applicable, is that the standard deviation of the instrument should not exceed one-half the standard deviation permissible in the overall measurement of the product". To illustrate this rule's soundness, he uses a right triangle with the short leg representing instrument variability (neglecting operator variability), the long leg representing process variability, and the hypotenuse, rep-

resenting overall variability being twice the length of the short leg. Values are then respectively proportional to 1, $\sqrt{3}$ and 2.

Further shortening of the short leg (testing variability) will not have much effect upon the length of the hypotenuse and hence will not pay off very much in reduced overall variability of product.

Testing precision, obviously important in certain types of scientific work or in testing valuable materials, is frequently a neglected subject in ordinary industrial work. This neglect can and does cost money. Money can be spent over-refining test methods. Costly tests of doubtful reliability can be run to give results not good enough to justify the cost of the test; or wrong decisions can be made from such results; or communication difficulties can occur from conflicting results obtained from such tests. Finally, wide testing variability in a quality characteristic which is tied to material cost may cost money by requiring unnecessarily high averages.

Possibly the two most important problems in testing precision are getting people to recognize that there is such a thing and that it can be measured. Unquestionably testing precision should be studied whenever a new test is set up. Usually it receives some general consideration under the heading of "reproducibility" but no attempt is made to get any quantitative measure of precision. Almost everyone knows of one or more sad occurrences that came about from unexpectedly large test (or sampling) variability (13). Why - under what circumstances do these things happen? Certainly much has been written about sampling and testing variability in recent chemical literature. Chemists are very skilled at finding causes of variability in certain kinds of tests. What kinds of tests cause trouble? Do they have any common characteristics? In spite of comments in the literature, in my experience at least, chemists do a lot of specifying workable sampling procedures, when they know that a sampling problem exists. When do they get into trouble?

One thing is sure, the attention paid to testing precision and sampling variability varies directly with the amount of "feedback" involved. If the test is to be rechecked within a short time, the effect will be anticipated and more precision built into the test, any inconsistencies that are found will be "fed back" to the original tester right away - and at an energy level corresponding to the dollar or other values involved. If the "feedback" is missing, delayed, at a low energy level, or contains the large amount of error (noise) of a highly unprecise test, then wide testing variability may go unnoticed for a long time.

This consideration of course, suggests the desirability of providing adequate "feedback"mechanisms for all testing procedures. Various types of mechanisms have been suggested, including standard samples and retesting of "blind" samples. In our company we have found that a "Quality Audit" program, in which monthly samples from our various plants around the world are sent to the Cambridge Research Laboratories and retested, has provided a workable "feedback" mechanism. It has actually given us a kind of communication network over which any quality problem is brought to light, considered and solved. In this respect it has pointed up very much the same thing that I have tried to get across in this paper. Quality control is something more than specifications, testing precision, statistics, human relations, processes and equipment. Quality control contains all of these but it may very well contain a

still bigger as yet undescribed philosophical concept and our success in working with quality may well depend upon our breadth of vision.

References Cited:

1. Ruth Benedict. Patterns of Culture. Houghton Mifflin Co. Boston, 1946, p. 2.

2. Hadley Cantril. The Why of Man's Experience. The MacMillan Co. New York, 1950.

3. Charles A. Bicking. "Quality Control in the Chemical Process Industries." Quality Control Handbook, edited by J. M. Juran. McGraw-Hill Book Co. New York, 1951, p. 565.

4. Preston C. Hammer. "Modified Limits for Control Charts". Parts 1 & 2. Industrial Quality Control, Vol. 2, No.3, November 1945 and Vol. 2, No. 6, May 1946.

5. A. R. Kamat. "Mean Successive Difference and its Ratio to the Root Mean Square". Biometrika, p. 116, June 1953.

6. Gyorgy Polya. How to Solve it. Princeton University Press, Princeton, N.J., 1948.

7. Alfred Korzybski. Science and Sanity - An Introduction to Non-Aristotelian Systems and General Semantics. The Science Press, Lancaster, Pa., 1st Edition 1933, p. 68.

8. Wendell Johnson. People in Quandaries. Harper & Brothers, New York, 1946, p. 53.

9. Jonathan N. Harris. "Product Line Earnings under Direct Costs", to be published.

10. Walter S. Shewhart. Statistical Methods from the Viewpoint of Quality Control. The Graduate School Department of Agriculture, Washington, D.C. 1939, p. 61.

11. Leslie E. Simon. "Contribution of Statistics to the Development and use of Purchasing Specifications and Standards of Quality". Fluid Mechanics and Statistical Methods in Engineering, University of Pennsylvania Press, Philadelphia, 1941, p.133.

12. Leslie E. Simon. "On the Relation of Instrumentation to Quality Control". Instruments, November 1946, Vol. 19, pp. 654-656.

13. A. H. Dodd. "The Assessment of Error". Food, March 1952, p.91.

THE USE OF CONTINUOUS SAMPLING IN AMMUNITION PROCUREMENT

R. L. Storer
Ordnance Ammunition Center

Introduction.

Continuous Sampling Inspection Plans are a relatively recent development in the field of inspection and quality control, but they are by no means new. The theoretical aspects of such plans have been covered by Dodge(1)(2)(3) and in a somewhat different form by Wald and Wolfowitz(4). Several detailed explanations of the operation of such plans are available. Our purpose here is not to retrace ground which has already been covered adequately, but to summarize experience gained in a wide-scale application of Continuous Sampling Plans (CSP) in the hope that this experience will be of use to you both in anticipating future developments in the field of government inspection and in applications for your own purposes as well.

The Plans we shall discuss are those developed by Harold Dodge in the references previously cited, adapted to the requirements of Ordnance inspection by personnel of the Ordnance Ammunition Center, an establishment of the U. S. Army located at Joliet, Illinois.

Ammunition Procurement.

The Ammunition Center functions as a sort of "home office" for the manufacture of all standard types of ammunition used by the Armed Services. The mission is accomplished in part with production facilities owned by the government, and in part by utilizing the output of a number of vendors and suppliers.

One type of government-owned plant operates much like an automobile assembly plant. Using metal parts and explosives manufactured at other locations, these loading plants add explosive elements to each of the parts and subassemblies, as required, and assembles the elements into a finished round or item of ammunition. Other government-owned plants manufacture explosives and certain metal components which for one reason or another are not available from the production capacity normally utilized by civilian industry.

Hundreds, at times thousands of vendors located in all parts of the United States are required to supply the loading and assembly plants. In addition, certain industrial firms have been given contracts for producing finished items which require no subsequent processing and are supplied by these plants direct to the using Services.

Our present inspection system requires that all material manufactured under an Ordnance contract be inspected prior to acceptance. Since a high proportion of the vendors are under contract to the government rather than to a prime contractor, it follows that there is a substantial government inspection organization involved in this enterprise, with inspectors working in nearly every state in the union.

The remarks which follow are based upon experience gained over a two-year period during which Continuous Sampling Inspection was progres-

sively applied, on a trial basis, to a wide variety of inspection problems. At the present time, CSP is in use on about 165 jobs in 8 loading plants and about 20 vendor's plants. The applications cover attribute inspection of clock-work fuzes, large bomb bodies, assembly of metal components for fuzes, subassemblies and assemblies of ammunition, filling shell with explosive charges, manufacture of cartridge cases, and a number of other manufacturing operations. The range is broad enough, we believe, to assure that the results are similar to those which would be experienced in the machining, metal working and assembly of metal components of a wide variety of civilian goods. The analogy would probably extend, also, to some aspects of the electronics and aircraft industries.

Typical Inspection Set-Up.

For economy reasons, most of the inspection done by the government in the field of ammunition procurement is source inspection—that is, inspection performed at the site of manufacture. The necessary attribute inspection is usually done in a series of stations or stages, each having associated with it a Classification of Defects (CD). Each stage is considered independently of the others so that final acceptance becomes a matter of passing a succession of inspection stages. The breakdown into stages is arranged primarily to avoid having to disassemble material in order to accomplish the required inspection, and has the side-effect of relating the inspection system rather closely to the manufacturing setup. The interdependence is so complete that in many cases the product must be inspected as it moves from one processing operation to the next.

Conventional Inspection System.

Prior to the installation of CSP, most inspection along conveyorized lines was done in accordance with the so-called "moving lot" provisions of MIL-STD-105A. The instructions provide, among other things,

(1) That the MIL-STD sample size be pro-rated over the expected lot size; i.e., if the expected lot size is 10,000 and the indicated sample size is 300, the inspector will usually collect his sample by taking three pieces from each 100 pieces passing the inspection station.

(2) That rejected material be returned to the contractor. Disposition of such material, if it is not government-owned, is a matter for the contractor's discretion. He usually screens the material if possible; otherwise it is reworked or scrapped. Government-owned material is screened or otherwise corrected.

(3) That material be allowed to pass the inspection station so long as the evidence for rejection is inconclusive.

This system led to a number of problems. The most serious of these, when it occurred, was the expense entailed in recovering rejected material. With product continuously in motion and being allowed to pass the station until proved rejectable, part or nearly all of the material involved in a lot might have passed through one or more subsequent processing operations by the time the reject number was reached. As a result, product would have to be unpacked, disassembled, removed from the production line, placed in a segregated area, or handled otherwise at considerable cost following the rejection decision. It might be thought that the expense involved in this situation is less a concern of the government than of the contractor. The presumption of such an argument is "Well, if the con-

tractor doesn't want the expense of unloading the product from a boxcar, he should keep the quality high enough to prevent rejection." The fallacy in this attitude is twofold:

(1) Often the government owns the material and is paying the entire operating expense on a direct reimbursement basis. In other words, some of the largest contracts are of the cost-plus-fixed-fee type.

(2) Even if the material were not government-owned, the increased costs occasioned by this kind of handling still have a tendency to be reflected in the price paid by our government for the product. Renegotiation clauses, waivers, prices for additional quantities of the material, etc., all afford the contractor a means to recoup such costs if they must be temporarily absorbed by him.

Another difficulty, less serious financially but more serious qualitywise, was the practice of "cutting-off" lots when the acceptance number was reached or approached. For instance, we normally consider a lot as being a day's production passing a particular inspection station. Suppose that by 1:00 P.M. of a certain day the acceptance number is reached. If another defect is found not only will all the material produced up to that time be rejected but also (since we expected the entire day's output to comprise the lot) a large number of pieces are rejected which are not yet even made. Much of the material, as we have indicated, may already have moved into subsequent processing operations and some of it may even be packed in the boxcar.

Since the inspection lot sizes are flexible, the manufacturer in such a situation usually calls an end to that particular lot at the point where the acceptance number is reached and starts a new lot and a new sample, thereby avoiding almost certain rejection. The disadvantages of this practice are evident. We do not, under such a setup, obtain the same sort of quality as we would get if we caused the entire day's production to be rejected and returned to the contractor for screening or correction. What is worse, the absence of direct and immediate action on the part of the government to reject the material leads to an unsatisfactory attitude toward quality on the part of both the inspector and the manufacturer. On the other hand, weighing the importance of some of the defects involved against the cost of unpacking, disassembling and reinspecting every piece in the lot, one cannot conscientiously go along with the drastic action a literal interpretation of the rules would require. As a result, the "lot cut-off" practice, far from being censured, eventually gained a certain measure of respectability and in certain instances is still accepted as standard operating practice.

Another difficulty encountered with the conventional inspection system is better classified as a nuisance than as a problem. It arises from the fact that lot sizes occasionally fall on the dividing line between MIL-STD sample size classes, so the inspector is faced with the necessity of picking up 100 to 300 extra samples near the end of the day if a lot runs a few hundred over an estimate; or he is faced with a decision as to which of his samples to declare surplus in case the lot packs out a few hundred short of the estimate. Aside from its nuisance value, this situation is not conducive to a good sample. In the first case, there is an excessive density of sampling near the end of the production run, and in the other case, some parts of the run are sampled to excess while others are not sampled at all.

Some of these difficulties can be corrected by accumulating or banking the material at the inspection station. However, this is not always satisfactory and more often than not is impossible. Banking of material increases inventory cost. A fuze manufacturer, for example, told us that elimination of banks at our CSP inspection stations caused a reduction of a quarter of a million dollars in his in-process inventory. Banking also increases the time required to convert raw or semi-finished material to a finished product. Product which must be allowed to sit in a bank while nothing is being done to it is, in a sense, product withheld from the production output. In time of emergency, it can have great importance. The increase in production cost and time required by inspection banks is not offset, in many cases, by any appreciable advantages. Banking does little to lessen the attractiveness of the lot cut-off feature, for instance, unless it is accompanied by a strict application of the static lot principle.

Essentials of a Continuous Sampling Plan.

Basically, a Continuous Sampling Plan of the type developed by Mr. Dodge operates in the following way. Inspection begins by screening the product 100% in the order in which it is produced (or as nearly so as possible), removing defects when found and replacing them with good pieces, and simultaneously keeping track of the number of pieces in each observed run of good pieces. When a run of "i" successive pieces free of the defect(s) for which inspection is being made has been observed, 100% inspection stops and sampling inspection begins. This period of 100% inspection is called the "qualification phase."

Sampling consists of inspecting one piece out of every 1/f pieces passing the inspection station. So long as this running sample contains no defects of the kind being inspected for, production is allowed to pass the station and is accepted. If a defect is found in the sample, however, 100% inspection is reinstituted immediately and continues until the process is again qualified by a run "i" successive defect-free pieces, whereupon sampling is resumed.

The parameters "f" and "i" are predetermined, and may be selected in such a way as to provide practically any desired degree of AOQL protection.

This procedure is known as CSP-1. There are two variants of this plan which Dodge has examined, known as CSP-2 and CSP-3. CSP-2 differs from CSP-1 in that return to 100% inspection does not take place immediately upon the finding of a single defect in the sampling phase, but only when the first defect is followed by another in less than "k" inspected units. CSP-3, in turn, is similar to CSP-2, except that the inspector examines the 4 units immediately following the first defect in the sampling phase. Depending on whether or not another defect is found among these four units, he reverts to 100% inspection or starts keeping count to determine if a defect occurs within "k" inspected units.

OAC Manual of Continuous Sampling Plans.

The Manual of Continuous Sampling Plans which is used on Ordnance Ammunition Center contracts was developed over a period of about 18 months from the experience previously mentioned. We started with a simple plan on one kind of operation, and as the diversity of applications grew, the character of our sampling plans also changed and eventually resulted in the series of plans reproduced here as Tables I and II.

TABLE I
Values of $\underline{i}$ for CSP-1 Plans by f, AQL and AOQL

AQL	f Values .005	.01	.02	.04	.08	AOQL
.015	2630	2210	1780	1380	1030	.12
.035	1970	1660	1340	1040	770	.16
.065	1370	1150	940	730	540	.23
.10	1170	980	800	620	460	.27
.15	880	740	600	470	350	.36
.25	530	450	370	290	210	.59
.40	380	320	260	210	150	.83
.65	300	250	200	160	120	1.08
1.00	240	200	160	130	100	1.35
1.50	150	120	100	75	60	2.20
2.50	100	85	70	55	40	3.09
4.00	65	55	42	34	25	4.96
6.50	43	36	29	22	16	7.24
10.00	28	24	19	15	12	10.70

TABLE II
Values of $\underline{i}$ and $\underline{k}$ for CSP-2 Plans by f, AQL and AOQL*

AQL	f Values .005	.01	.02	.04	.08	AOQL
.015	3200	2650	2200	1750	1350	.12
.035	2400	2000	1650	1320	1000	.16
.065	1660	1380	1150	920	710	.23
.10	1410	1180	980	780	600	.27
.15	1060	880	730	590	450	.36
.25	640	540	450	370	280	.59
.40	460	380	320	260	200	.83
.65	360	290	250	200	160	1.08
1.00	290	230	200	160	130	1.35
1.50	180	150	120	95	75	2.20
2.50	130	110	85	65	55	3.09
4.00	75	65	55	42	32	4.96
6.50	55	42	35	28	22	7.24
10.00	33	27	23	19	14	10.70

* $\underline{k}$ is equal to $\underline{i}$.

These tables utilize values of f which require sampling frequencies ranging from one piece out of each 200 to eight pieces out of each 100 passing the inspection station. The qualifying periods vary accordingly, ranging from 12 pieces to over 3,000 and the associated AOQL's range from .12% to 10.7%. It should be noted that AQL is used only as a label, in our practice, and does not imply that the characteristic curve passes through any particular point. In the two tables we are discussing, the AQL is included primarily as an index to the MIL-STD-105A plan which would yield about the same AOQL as the CSP plan if it were used on an AOQL basis. In most instances, the conventional inspection system actually does have AOQL connotations, for the customary action taken by the contractor upon the return of a rejected lot is to screen the lot under the supervision of an Ordnance inspector, removing all defects of the kind causing rejection.

In addition to the wide range of i, f and AOQL values, plans developed at OAC also provide for the use of either CSP-1 or CSP-2. Each feature of the plans we now have is the result of a problem or situation in the field which required an expansion, addition or change to the simple plan with which we started in 1952. We therefore expect that the set of plans will have wide and successful use in the future.

Effect of CSP on Quality Consciousness.

Initially, we investigated the use of CSP in the hope of eliminating the expensive recall provisions attendant upon the conventional inspection system and concurrently of reducing the amount of material which had to be banked near inspection stations. When these banks consist of explosives, they become safety hazards as well as production and fiscal problems. There is no question but that the CSP plans either eliminate or materially reduce troubles of this sort.

The most important benefit, however, was unexpected. Under the conventional inspection plan, as I have previously mentioned, rejection did not follow until sometime after the finding of the first defect. The MIL-STD-105A plans most often used have rejection numbers as large as 15. Several hours' production is ordinarily required to generate enough defects to cause rejection. In view of the "lot cut-off" feature discussed earlier, rejection may not follow a run of quality trouble, except under very difficult circumstances. By contrast, CSP-1 causes immediate rejection of the process and use of screening as an expedient means of quality correction upon finding of the first defect. CSP-2 is a little weaker in this respect, delaying corrective action until a second defect is found too close on the heels of the first. In either case, however, action is sufficiently direct, and the severity of the consequences is such, that the need for attention to quality becomes quite evident at all levels from operator to production superintendent. Inspectors and production supervisors have both remarked on the increase in quality consciousness as an outstanding by-product of CSP. These comments, be it noted, are not isolated but have arisen independently from nearly all users of the plans.

Concern was at first expressed that CSP tended to venture into the field of process control and might come to be used as a replacement for controls which the manufacturer ought to maintain himself. So far, the observed effect has been the opposite. Following the introduction of Continuous Sampling Plans, quality control programs have sprung up and prospered in plants where there had been literally no process control

system before. It is our opinion that this is a natural consequence of the use of CSP. Any attempt to rely on the government's use of CSP, alone, as a means of controlling outgoing quality is almost sure to fail if the quality is uniformly bad or very erratic. The reason is that a constant load of 100% inspection for all defects is a possible, and in some cases even a probable sequel to poor quality. Worse yet, there may be a rapid alternation between periods of 100% inspection and sampling inspection. The manufacturer has to supply the manpower necessary for 100% inspection phases, and the 100% inspection itself tends to slow down the rate of production on the line. Hence if the absence of adequate process controls led to a deterioration of quality, the manufacturer would soon be apprised of the situation in terms which are very real to him—namely increased labor costs and lowered production rate. One or two experiences of this kind are enough to convince him that preventive measures are in order.

Sampling Frequency.

Determination of the f-value to use in a given situation is a matter we have been unable to reduce to a simple formula. On one hand, the relative amount of inspection should be kept as small as possible to minimize costs, which means that f should be small. On the other hand, if f is too small, short segments of bad quality may be overlooked because they are not sampled.

A natural first solution to this problem is to specify a larger f for the important defects than for the unimportant defects. In our first plan, for example, we provided a 4% frequency for Critical defects and a 2% frequency for Major defects. We found, however, that this practice creates a difficult situation if both kinds of defects are inspected for at any given station. It requires that the inspector take 4 from each 100 pieces produced as a sample for Critical defects and of these 4 that he choose 2 to be examined for Major defects. Suppose that while inspecting for Criticals, or while in the process of drawing the sample, he discovers a Major on one piece. Should he then call that piece a part of the Major sample or not? Or worse, suppose he does not notice the Major until he has picked 2 of the other 3 samples as representing a sample for Major defects and has found these 2 to be free of Major defects. In view of the potential for differences of opinion in such situations, we have adopted the practice of using a single value of f for all defects at a station.

Some of the methods we have used in determining f are as follows. It should not be inferred that these constitute a complete set of rules. These statements are merely considerations which we have followed with reasonably good results so far.

(1) If the material is simultaneously processed in batches, we use a value of f which will provide a sample of at least one piece from each batch.

(2) If the inspection station is concerned with an assembly operation or a relatively stable and simple series of machine operations, we usually use $f = .01$ provided that the production rate can be expected to fall somewhere in the range of 5,000 to 20,000 pieces per day.

(3) If the production rate is very low, say less than 500 pieces per day, we usually use $f = .08$.

(4) If the production rate is extremely high, say 50,000 per day or more, an f of .005 is indicated in most cases.

(5) For situations intermediate to those previously named, we select a frequency which seems to balance the factors of minimum amount of inspection, reasonable amount of information, and adequate protection against spotty quality. While it has, on occasion, been advisable to consider each manufacturer as a separate problem in determining the value of f, we have found that the sampling frequency can more often be specified by item or by the type or nature of the operation. For example, shell filling is, for the most part, a batch process, the batches being either 25, 36, 54 or 80 pieces, depending upon the size of the shell. Hence we specify 4% for 25 - 36 shell batches, 2% for 54-shell batches and 1% for 84-shell batches, or 1% in the case of continuous pouring in lieu of batch processing. Large bombs are usually produced at a very slow rate and an 8% value of f therefore fits most bomb manufacturing operations. Almost all conveyorized assembly operations may be satisfactorily handled by a frequency of 1%. As a result of "cut and try" methods of this kind, we should eventually be able to provide fairly simple rules for inspection procedure writers to follow in selecting the value of f to be specified. If, in the meantime, someone is able to present a straightforward method for selecting the optimum value of f, we shall all be much better off.

Amount of Inspection as Compared with Conventional System.

Table III summarizes experience with the conventional and CSP plans on a shell-filling inspection at 9 plants over a period of several months. The notations "CSP" and "105A" in the right hand column of the table refer to continuous sampling plans and moving-lot inspection by MIL-STD-105A, respectively. Most of the CSP inspection involved was done under plans designed to approximately equal the conventional plan in relative amount of material inspected on a sampling basis. Even so, it is evident that the use of CSP caused an important reduction in work load at some plants. On the other hand, it caused notable increases in screening in others. Considering the overall picture, the total amount of sample inspection remained approximately the same on CSP as under the conventional plan. Obviously, this circumstance arises because the CSP plan was so designed. It is equally clear that CSP could be made to show either a reduction or an increase in the relative amount of material inspected on a sampling basis depending upon the relation between f and the corresponding fraction m/M of a MIL-STD plan.

Apparently, substantial reductions of the total amount of inspection required under the conventional plan can be made without much, if any, sacrifice in outgoing quality. For instance, we recently investigated the outgoing quality in a very difficult situation. The operation involved, known as melt-pouring, is sometimes tricky to control. Quality can change abruptly without anyone knowing very precisely when or why the change occurred. A shell-filling job developed serious quality trouble twice during the filling of about 35,000 shell. A 2% CSP plan was in use at the time this difficulty arose. The effect of the CSP plan on the outgoing quality can be judged by the following table which was reconstructed from inspection results during the manufacture of the material and by inspecting all material passed by the inspector.

TABLE III
Summary of Inspection Results

Facility	No. of Lots Presented	Total No. of Units Presented (Thousands)	Avg. Sampling Insp. Frequency (%)	Process Average, % Critical	Process Average, % Major	100% Inspection (% of Total Presented)	Sampling Scheme
Iowa Ordnance Plt.	56	3,591	3.68	.000	.020	0.22	CSP
	295	5,400	5.48	.004	.187	2.03	105A
Joliet Arsenal	355	5,245	4.87	.001	.009	0.04	105A
Kansas Ordnance Plt.	443	5,117	3.23	.001	.074	1.35	CSP
	127	1,222	3.71	.000	.066	0.33	105A
Kingsbury Ord. Plant	17	78	3.29	.000	.000	0.00	CSP
	220	825	5.19	.000	.070	3.61	105A
Lone Star Ordnance Plt.	82	1,843	3.31	.026	.310	3.87	CSP
	341	5,897	3.26	.004	.061	0.86	105A
Louisiana Ord. Plant	249	2,974	3.54	.001	.009	0.19	CSP
	355	5,819	3.55	.002	.063	2.12	105A
Milan Arsenal	50	1,000	4.48	.002	.011	1.42	105A
Nebraska Ord. Plant	176	5,248	1.77	.019	.050	1.29	CSP
Ravenna Arsenal	254	3,073	4.91	.001	.247	3.16	CSP
	136	1,130	6.73	.085	.291	1.70	105A
Totals	1277	21,924	3.24	.007	.108	1.45	CSP
	1879	26,538	4.37	.008	.096	1.33	105A

TABLE IV
Outgoing Quality Obtained by Use of CSP
in a Situation of Poor Quality

Quantity Mfg'd	% Inspected by Sample	Total % Inspected during Production	% Defective at Time of Presentation	% Defective of Accepted Material	AOQL
34,381	1.82	30.97	Cr. .84 Ma. .38	Cr. .08 Ma. .19	Cr. .12 Ma. 1.09

The 31% total of product inspected during production is a reflection of the extent and duration of trouble encountered and will, of course, vary with the frequency with which the defects occur. Thus, at the expense of sampling less than 2% of the product (as opposed to more than 3% under the conventional plan) and of detailing 30% of the output, the fraction defective of Criticals was detected and reduced from .84% to .08% in the accepted material. This is about the same result as would be obtained under the conventional MIL-STD-105A approach by 100% reinspection at 90% efficiency. The cost of doing the job by CSP is estimated in this case to have been about 10% of the cost by conventional means.

Application of CSP.

CSP seems to appeal intuitively to inspector and producer alike, so training in the use of the technique is not difficult. After the first pilot application, we called all District and Plant inspection administrators together for a training session in which we discussed the basic ideas of the plans rather thoroughly. These administrators, in turn, have made it a practice to discuss CSP very thoroughly, both pro and con, with the producer. In talks with the producer, the intimate relationship between quality, rate of production and cost is stressed. Special pains are taken to insure that the producer fully understands that a sequence of defects will reduce his rate of output to that of the screening crew, and that the screening must be done properly. This approach has been very effective in creating a team spirit with respect to CSP, has assured thorough and considered trial and has minimized the confusion which always accompanies a change in method.

It is sometimes said by persons considering its use that CSP should not be used except where the quality history is good. Normally, we make the same recommendation, but for a very different reason. The person to whom CSP is new is likely to want to try it only on stable high-quality operations because, he says, he is not satisfied that it gives him "equal assurance." One who has used it is not likely to be worried about the "assurance" (whatever that may be) but is worried at the effect on the production rate if CSP is applied on an erratic or poorly controlled line.

Inspection stations set up for the conventional plan and for inspection on the basis of the defective may need to be re-organized under CSP. Defect lists may be too long for rapid inspection. Or, some disassembly may be necessary under the conventional plan. Usually, we find, the inspection can be done with minimum manpower and maximum effectiveness by locating the necessary inspection as close as possible to the place where the defect originates. This practice leads to some troubles of AQL re-

adjustment and similarity of inspection reporting if the defective continues to be the basis of judgment. Accordingly, we usually provide for judgment by defect under CSP. This is simple and straightforward, eliminates one bookkeeping operation, prevents such difficulties as having to clean up all Minors because of the occurrence of one kind of Minor in a sample, and greatly simplifies the screening operations.

Contractual Relationships.

Several contractors have commented that relations between the government inspector and the contractor have improved noticeably following the introduction of CSP. They say there is much less friction; that the inspector's decisions, while no more popular than before in case of a rejection, seem to be more soundly based and more readily accepted. It is possible that the inspector may be more keenly aware of his responsibility, in view of the drastic effect his decision may have. It is also possible that the contractor, freed from the worry of a rejection under the difficult situation of having to recall and reinspect product, takes a more charitable view of government inspectors. Whatever the reason, this mutual increase in respect has been most heart warming.

Scope of Future Applications.

CSP is, I think, going to play an increasingly important part in the design of inspection plans and systems because it is consistent with a sound fundamental of inspection planning. Specifically, an inspection plan, including especially its sampling phases, will be useful and efficient in direct proportion as it is designed to fit the needs and limitations of a particular production and inspection situation. MIL-STD-105A is so designed that it works best when the material to be inspected is a finished and static mass, from which samples may be drawn and inspected, with decision following upon completion of inspection of the entire sample. For Receiving Inspection or any other attribute inspection of static material, it is difficult to equal.

Conversely, it should occasion little surprise that the same Standard leaves something to be desired when applied in a climate of multistage inspection, moving product, and what amounts to decision piece-by-piece. Nor is it very remarkable that the Dodge Continuous Sampling Inspection Plans, designed for just such a situation, should fill the need more effectively and efficiently.

The probability of successfully combining the ideas in NAVORD-OSTD-81 with those in the OAC Manual to form a standard for attribute inspection of moving product suitable for wide application in the Department of Defense is, I think, very good. Standardization is feasible if it be kept in mind that many people will require the plans for many situations; that therefore standardization cannot be synonymous with rigidity of procedure; that a good standard must really be somewhat like a set of Standard Gage Blocks--a set of known elements from which each planner can fashion a setup to suit his particular needs.

I believe, also, that CSP has great potential in industrial usage as a final inspection device and as a control on assembly or other manual operations where quality is largely a function of the individual's attention to his work.

Summary.

If it had done nothing else, this experience with CSP would have been worthwhile because it probably means the end, or the beginning of the end, of an era of thinking in terms of an all-purpose set of Sampling Tables. The advantages to be gained by accommodating the inspection to the production situation, rather than vice versa, are obvious. In addition, the evidence seems pretty compelling toward the following conclusions:

(1) Properly applied, CSP encourages the growth of quality consciousness and, when used for acceptance purposes, encourages the use of effective control during processing.

(2) As a replacement for a plan based upon the use of MIL-STD-105A on moving product, CSP has saved and will save substantial sums of money by eliminating the need for recalling material which has already passed the inspection station and by forcing corrective action on product yet to be made.

(3) CSP has demonstrated that under certain circumstances it is not necessary to inspect large fractions of the output, as compelled by the smaller lot sizes of MIL-STD-105A; on the other hand, it has cast some doubt upon the wisdom of a casual use of the very large lot sizes of MIL-STD-105A, particularly under reduced sampling.

REFERENCES

(1) Dodge, Harold F., "A Sampling Inspection Plan for Continuous Production," Annals of Mathematical Statistics, XIV, 1943, pp. 264-279.

(2) Dodge, Harold F., "Sampling Plans for Continuous Production," Industrial Quality Control, IV, No. 3, 1947, pp. 5-9.

(3) Dodge, Harold F. and Torrey, Mary N., "Additional Continuous Sampling Plans," Industrial Quality Control, VII, No. 5, 1951, pp. 7-12.

(4) Wald, A. and Wolfowitz, J., "Sampling Inspection Plans for Continuous Production Which Insure a Prescribed Limit on Outgoing Quality," Annals of Mathematical Statistics, XVI, 1945, pp. 30-49.

VISUAL AIDS TELL THE QUALITY CONTROL STORY

Paul C. Clifford
State Teachers College, Montclair, N. J.

"The fundamental concepts of Statistical Quality Control are very simple-- or so it seems to those who believe they understand them. But in trying to explain them to any group of people, the common experience is one of difficulty in getting the basic concepts understood". (1) I might emphasize that this difficulty is one encountered by college professors, quality control managers, shop foremen and line inspectors.

My purpose today is not to discuss the basic concepts of Quality Control, but, assuming that there might be some reasonable agreement as to what these are, to explore the problem of communicating these concepts to groups that are not familiar with them. Frequently the difficulty may be due to our effort to explain these concepts verbally, neglecting the picture, gadget or model that would replace a thousand words. To emphasize the value of models as opposed to words, let me remind you of what happened in Korea when Marilyn Monroe appeared.

Next let me suggest that Quality Control has many aspects, as is obvious from the breadth of papers presented at this convention. To some it is applied statistics, to others it is a management method, while there will be some that claim that it is a body of empirical tools. All will agree however that there are psychological aspects that cannot be neglected. I think that every discussion on the management aspects of control chart work has indicated the effect on the production worker of having a quality record that he understands. The psychological reaction of the audience to the visual aid is the important item to keep in mind, and not the accuracy of the particular gadget. With this much background let me proceed to a brief discussion of some visual aids.

The first basic concept in quality control is that of a distribution resulting from a stable cause system. Probably the best demonstration of this is an actual production unit, such as an automatic lathe. Several In-Plant training programs have used successive items of such a production unit as the introduction to the concept of a distribution. This is excellent for the particular industry, but may have less application for a general group.

The best gadget to demonstrate the concept of a distribution for quality control purposes is the Quincunx, first described by Francis Galton in 1889. Several of these are available commercially, and it is my firm recommendation that you buy one rather than make your own. This recommendation is based on a sample of one- the one I made. Sources where various models and gadgets may be secured will be listed at the end of this paper. In the Quincunx I have here, the beads fall from an outlet that can be adjusted to any desired lateral position. They then pass through some ten rows of nails and fall into slots. There is a retaining rod a few inches below the top of these slots, so that a sub-sample of any desired size can be collected. This retaining rod can be manipulated with a plunger, so that the sub-samples can be accumulated on the lower half of the quincunx. Figure 1 shows a quincunx along with some dice, a bowl of chips and a box of beads. These will be discussed later.

Let me suggest some sales talk to go with this demonstration. Every industry complains about the large number of variables that affect the

quality of their product, without realizing that this is the very reason they need statistical quality control. If we consider one product- say paper board for folding cartons- and list the variables that affect such a characteristic as moisture we can easily make up a list of ten or more factors. Some of these factors will be much more important than others, but also the ability to control these factors is usually higher. Frequently the effect of each factor is of about the same magnitude.

Think of each factor as being represented by a row of nails, the first row representing the usual variation in one type of raw material, the next row a preliminary step in the process, until the last row may represent the error in measurement of the finished product. The resulting distribution is caused by a constant system (the ten rows of nails) of chance causes (which way will they bounce?) Nothing is being changed as unit after unit is run through this process, but we still get a distribution of this quality characteristic.

Let us now introduce an assignable cause by shifting the bead outlet. Depending on the amount of this shift, which may be as much as 2's, will be the number of observations needed before we can detect the shift from the resulting readings. If we shift only one standard deviation, we would expect one of the first fifty beads to fall outside the limits of a large sample taken at the original position. We conclude that our ability to catch a shift of this amount by observing individual results is very poor.

Assignable causes can be introduced by shifting the bead outlet, which is psychologically the easiest to translate into production terms, by changing the slope of the board, or by tilting the board. I have seen some where the whole nail pattern could be changed, both as an entire unit and row by row. Increasing the slope of the board tends to give increased variability (What happens when you speed up a production line?) Gradually moving the bead outlet tends to give a flat topped distribution (What happens when you have trends, such as tool wear?) Tilting the board tends to give a skewed distribution. (What happens when you try to keep to the top limit on a machine cut?)

It should be obvious that such a gadget can be used to construct charts that illustrate a system in control, as well as to illustrate how the presence of an assignable cause is indicated. There are many other situations that the ingenious expositor will be able to illustrate with this gadget. That will be true of all the aids that I will mention, and I will try to point out only the major applications.

Dice are very useful for a variety of purposes. Moreover they seem to have an intrinsic appeal not possessed by any other visual aid. I have a small collection here that represent an investment of some three or four dollars. As you will see, it is an investment that could return rich dividends- not without a considerable risk however. One pair is a fair pair - honest dice. Or at least they are as honest as you could expect for fifty cents. Next we have the type called misspots (pronounced mis-spots or miss-pots?) of which I have six. Here the individual die is not numbered one to six, but certain numbers are repeated.

Die A	All 5's
Die B	Three 2's, Three 6's
Die C	Two 1's, Two 5's, Two 6's
Die D	Two 3's, Two 4's, Two 5's

Die E	Two 1's, Two 3's, Two 5's
Die F	Two 2's, Two 4's, Two 6's

Let me admit here that these "visual aids" were designed for a different type of educational usage. However with judicious handling they can be used to illustrate typical situations in quality control. The first pair which result in 7's and 11's only illustrate the obvious or catastrophic defect where little or no statistical analysis is required. A broken die in a punch press may be an illustration. Dice C to F can be combined in ways that give somewhat less unusual results. Such effects as a shift in average, increase and decrease in variability, unusual sequences as all odd or all even can be obtained. Here the proper psychological setting is to explain that we must attempt to determine control (honest dice) or lack of control (misspots) from the results, and not from an examination of the dice.

If these seem too obvious, you might prefer a set of dice which are correctly marked, but which have unequal probability for the different faces. The set I have are called "shapes" I believe, and the bias could be detected by calipering. I understand that dice can be made with this same bias by other methods, harder to detect, but for a greater price. The set I have have probabilities approximately:

	Probability of					
	1	2	3	4	5	6
Die X	.22	.14	.14	.14	.14	.22
Die Y	.22	.14	.14	.14	.14	.22
Die Z	.14	.14	.22	.22	.14	.22

Once the subject is psychologically prepared for the misspots, hand him these shapes and he will almost always check for misspots. I have had students roll these dice many times and decide with great assurance that these were honest dice. This illustrates one mathematical point and one psychological one. How long would it take you to detect a change from 16% to 22% defective? Secondly how often are we checking for the type of defect that occurred last time and failing to consider other important defects that could occur?

While I am on the subject of dice let me mention that they may be used rather effectively to illustrate the addition of tolerances that individually have a rectangular distribution. For n honest dice, the sum has an expected value of 3.5 n with an expected variance of 35n/12. For n greater than 4 the distribution of the sum is close to normal. Thus for eight dice the three sigma limits for the sum is 14 to 42 as compared to a total range of 8 to 48.

Next let me mention the control board. Here a distribution is strung out on wires, so that the beads may be arranged in the order of observation. The number of wires and the size of the beads used can be varied for individual purposes. The distribution I have used (Figure 2) is:

X	-4	-3	-2	-1	0	1	2	3	4
f	1	2	4	8	10	8	4	2	1

The easiest way to set up a chance or controlled set of observations is to make up a set of chips with numbers and frequencies corresponding to these, draw chips and set the beads correspondingly. Figure 3 shows such a control pattern. Figure 4 on the other hand shows a definite

trend in addition to a "wild shot" on the second sample. This illustrates the need for considering the observations in some rationally sub-grouped manner as opposed to compiling all observations into a distribution. Also it illustrates the difference between capability and achievement. This board is similar to a much larger Control Board used by A. G. Klock and C. W. Carter of Bigelow-Sanford Carpet Company, described in the September 1950 issue of Industrial Quality Control.

At this point let me digress from the list of visual aids to discuss one aspect of their use. In the past I frequently have been asked to talk on "Basic Concepts of Quality Control" for various groups. I go to one of these meetings loaded down with models, gadgets, case histories etc. However you have to be careful not to use them when they are not needed. On the other hand you never know when someone will ask a question the answer to which could be quickly illustrated by the right gadget.

For example, here is a minor item (Fig 5). It consists of a grooved base in which can be inserted any one of a number of distribution curves. In front of these grooves are holes drilled to hold 1/8 inch plastic rods. I have beads of various sizes that can be slipped over the rods. By properly combining beads and curves I believe that I can quickly illustrate the idea of a histogram and its underlying population for most all n's from 40 to 200. Very seldom needed, but most illuminating for the person who just cannot understand the verbal distinction between statistic and parameter.

Next let me discuss what I consider the best model to show the power of a control chart for averages (or medians or mid-ranges) as opposed to inspection of individual items. This is simply a model of a distribution with snap fasteners attached so that a smaller template showing the distribution of the average (median or mid-range) of samples of size n can be attached. The distribution need not be normal, although that is the most useful one. Figure 6 shows such a model. Rods may be inserted at desired points along the base line. I have used red rods to indicate the specification limits, and green rods to represent the control limits. If we shift this distribution two sigma to the right we get Figure 7. This shows that under the conditions given the chance of observing such a shift is about .03 if we look at individual observations outside the specification limits, but is about .95 if we look at the average of 5 pieces. This illustrates the old adage "Keep your eye on the process, not on individual results". Figure 8 shows the one model with modified control limits shown by yellow rods. Thus this one model can be used to clear up the concepts control limits, modified control limits, process capabilities and specifications. If the templates of averages of samples of size 5 are replaced first by templates of samples of size 2, and then by templates of samples of size 20, we begin to appreciate the importance of sample size and the problems of under control and over control.

Figure 9 shows the templates adapted to a slightly different problem-that of explaining the different ideas in accepting sampling by variables. Here the optimum lot quality is one with average 25, upper and lower tolerance lots have averages of 29 and 21, sigma is known as 2.5 and attaching templates for different lot size <u>n</u> we find experimentally that samples of 20 give excellent discrimination.

In addition to just slippage of the mean we need to be able to illustrate the effect of an increase in dispersion. To do this with complete

accuracy is not possible. We must remember that a visual aid is meant to convey the correct impression but need not be technically correct itself. Figure 10 shows a distribution with the inner rods set at R and the outer lines set at ± 3s which are taken as specification limits. Figure 11 shows the effect of an increase of variability of approximately 50%. The average range has increased 50% and the product now has about 5% outside specifications. It is obvious that any slippage of the mean with this increase in variability will result in a large percent outside the specification limits.

For construction of a control chart the use of numbered chips for a theoretical universe is well known and is discussed in most any text on the subject. I have found that the use of a specific set may be much more convincing. The set I have here consists of 200 numbered chips that correspond exactly to a series of 200 observations on a chemical process. When the data were plotted in the order of observation (sub-groups of five) both the average and range showed lack of control. When random samples of five are drawn from the same identical set of observations, they naturally show control. This is one of the most convincing demonstrations I know that pure chance will not show the same condition that we ascribe to lack of control. It is especially convincing if the data concerns a practical problem that the group is familiar with.

Frequently a set of variables data may be classified in two or more categories. The methods of analysis of variance can be used, but the concepts involved are not easily communicable to the novice. As an example consider a set of readings from a calendering process, where position on the calendar as well as time may be important. For this type of problem a special model may be constructed. Figure 12 shows a model that represents the variation in caliper during one shift's production. The height of each plastic rod represents the caliper (.020 being the base plane). By viewing the model both frontwards and sideways the novice can see both the cross machine pattern as well as the variation in time. This particular board has 7 columns and 26 rows, so that almost any problem of this type can be set up quickly once a supply of precut rods for the deviations from a datum plane are available.

Sometimes the observations in each column can be better represented by a frequency distribution. For the usual correlation or regression problem this is frequently the case. Figure 13 shows a correlation board consisting of eleven columns, each with 17 cells. The number of beads on a rod represents the number of observations in a cell. By shifting the grooved columns we can easily illustrate such concepts as linear regression, standard error of estimates. Figure 14 shows the same bivariate distribution as before, with the regression removed. This model can also be used for two way analysis of variance.

The last two models illustrate some of the difficulties involved in the visual approach. It will take the instructor at least one hour to set up the model to illustrate a particular problem. Which puts the question - What is the value of the model makers time as compared to the model viewers time?

As a final item concerning variables let me suggest the set of building blocks to show the combination of tolerances. This consists of five different set of blocks, each set of 20 forming a distribution that is approximately normal. The frequencies I used from the lowest cell to the highest in each set are *1,1,2,3,6,3,2,1,1. One set has cell intervals

of 2/16, two have intervals of 3/16 and two have intervals of 4/16. In addition each set has a different average length. The largest blocks from each set are assembled, as well as the smallest blocks, and the respective total length marked on a board. Then these blocks are put back in their respective sets and random assemblies of one block from each set are made. After a reasonable number of these have been assembled, three sigma limits for the assembly are computed and found to be approximately one half the total spread of the largest and smallest assembly possible.

I have left the question of attributes until the last. There are various gadgets to illustrate attribute sampling but none that I have seen are better than the box of beads and sampling paddles. I would suggest that sampling paddles holding 5, 25 and 100 beads each be made. There is no logical reason why one could not explain that one row of the large paddle would represent the sample of five and thus use only one paddle. It may be that there are psychological reasons for having a small paddle, where the rest of the sample cannot be observed. The same question arises as to whether the container should be transparent or opaque. Personally I use three or more distinct boxes of identical opaque material to try to establish conditions similar to those we encounter in actual sampling. Each box consists of 1000 beads in the following proportions:

	White	Red	Blue	Yellow
Box A	94%	4%	1%	1%
Box B	89.5%	8%	2%	0.5%
Box C	84.8%	12%	3%	0.2%

Since many more colors are available, boxes of greater complexity can easily be arranged. These given will cover pretty well the range from .2% to 15% defective, and can be used to illustrate both single and multiple classification of defects as well as the demerit approach.

A word in closing. Let me emphasize again the need for a psychological approach to any evaluation of these gadgets. When we use them we are trying to convey an idea, not to give a technically correct statement. As an example I have received criticism on the model with the template of averages to the effect that each distribution should be the same size, with the result that the template for averages should be n times as high as that for individuals. There is a technical argument here that is correct. However the impression that the non technical audience gets would be confused and probably wrong. To them it is obvious that there are fewer averages than individuals. Why confuse them with an argument that is not concerned with the fundamental idea of what control means?

References:

(1) Ott, Ellis R. and Clifford Paul C., Basic Concepts of Statistical Quality Control Proceedings of Rutgers Conference on Quality Control September 15, 1951

Sources of Models

The Quincunx used and described is made by Walter P. Koechel of Tung-Sol Electric Company, Bloomfield, N. J. and is available commercially from him. Another Quincunx illustrated on the March 1950 issue of Industrial Quality Control is made by R. E. Wagenhals of the Timken Roller Bearing Company, Canton, Ohio.

Dice of the type discussed are available through many magic supply

houses. The K. C. Card Company Chicago, Illinois lists hundreds of different types.

The rest of the models used were made by Ray R. Lilly, Midvale, N. J. and are available from him.

Mr. Koechel also has a control board similar to that shown as well as a board similar to that in Figure 8.

There may be other sources where these are obtainable. No exhaustive search has been attempted.

Figure 1.

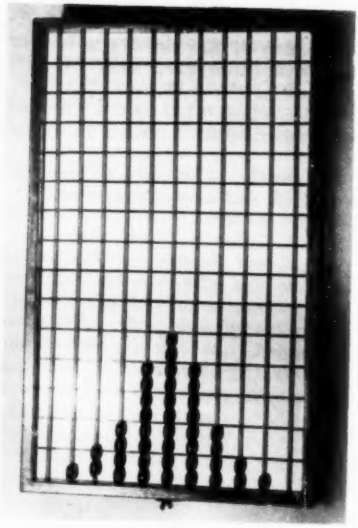


Figure 2

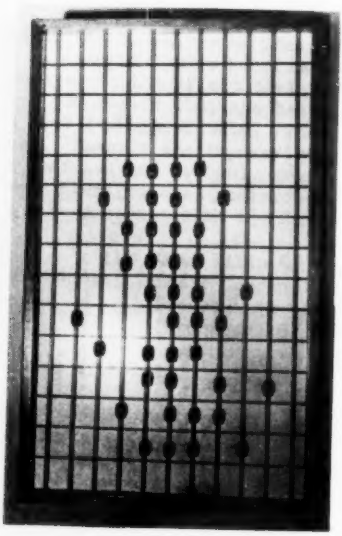


Figure 3

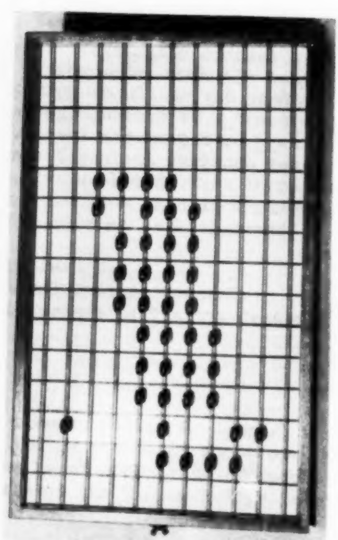


Figure 4

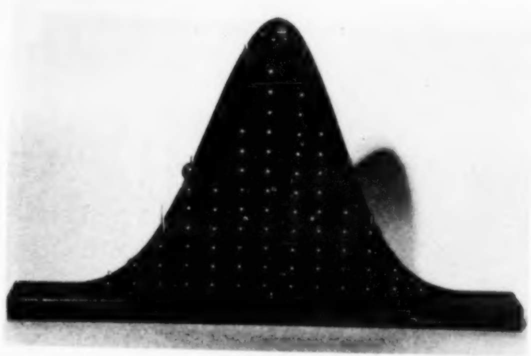


Figure 5

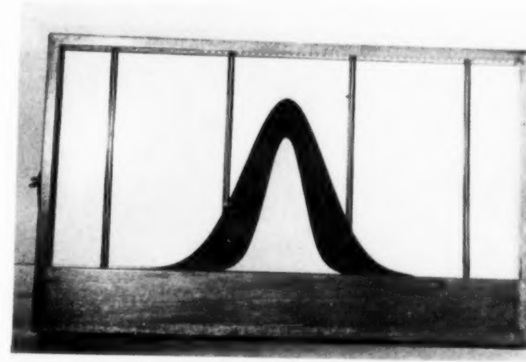


Figure 6

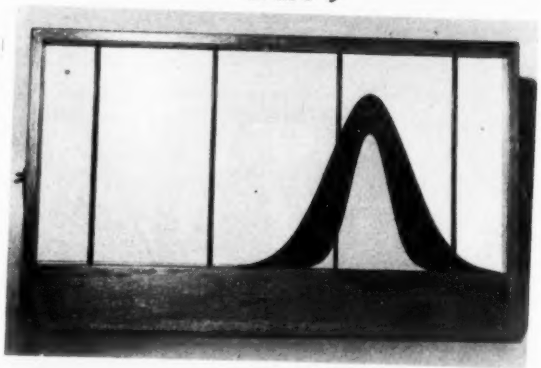


Figure 7

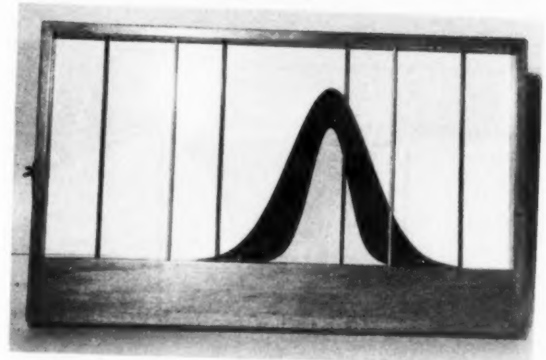


Figure 8

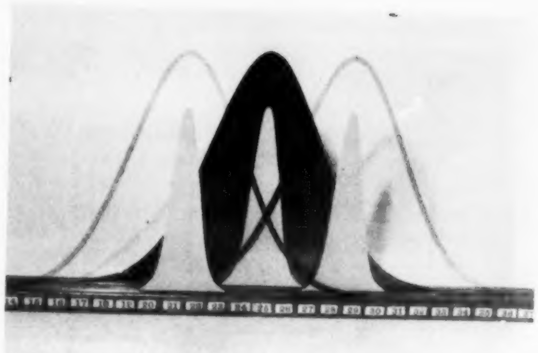


Figure 9

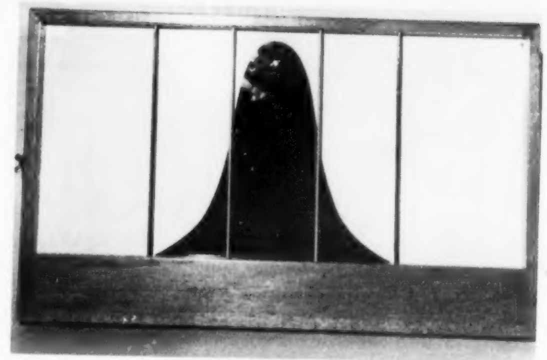


Figure 10

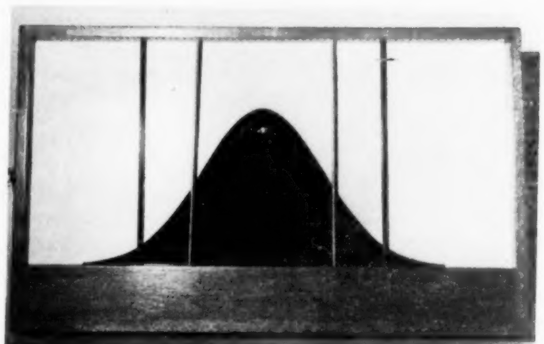


Figure 11

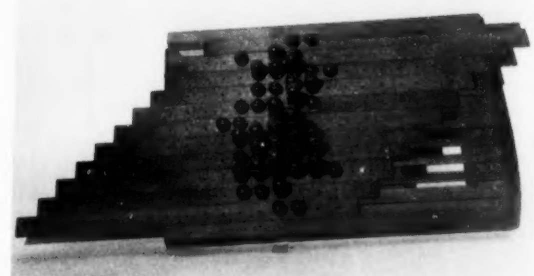


Figure 13

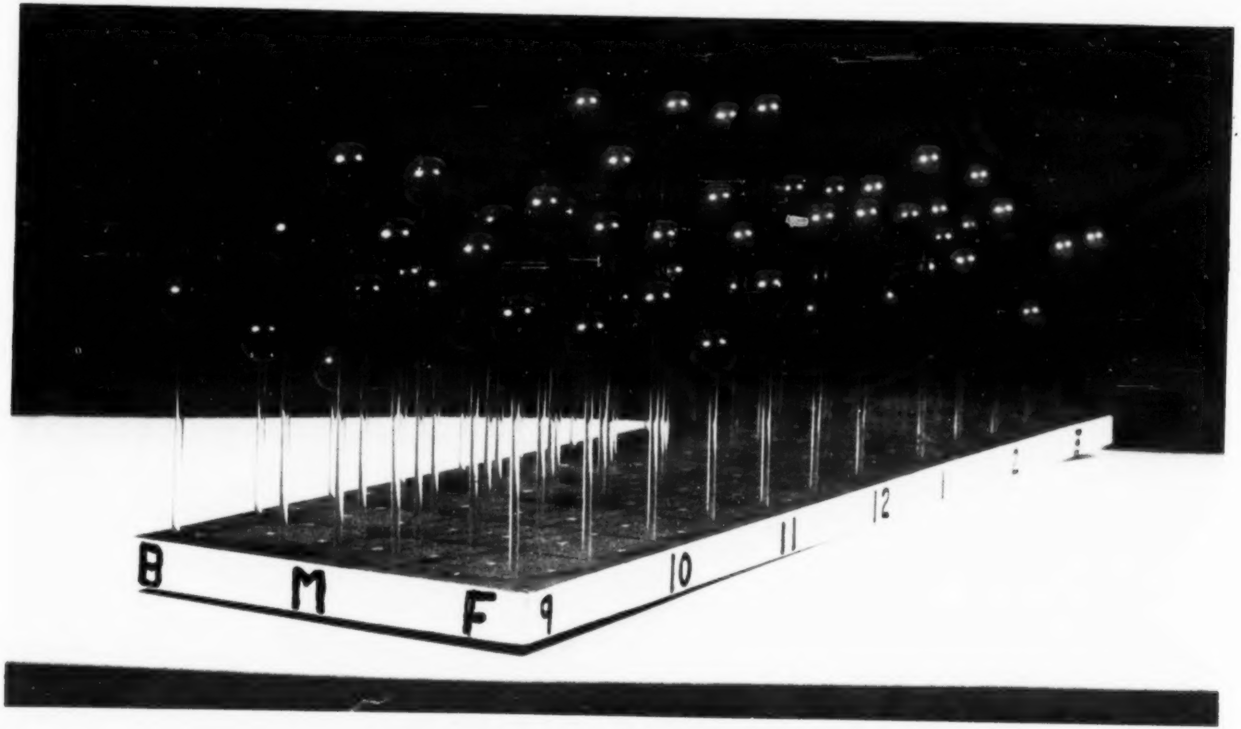


Figure 12

FUNDAMENTALS OF ANALYSIS OF VARIANCE

Charles R. Hicks
Purdue University

In the last few years there has been considerable development in our thinking on the technique of analysis of variance. The classical approach to the problem of which many of you are familiar and which is found in all but the latest textbooks is now actually a special case of an extended theory of experimental designs based on the use of mathematical models. It is my purpose to review the classical technique of analysis of variance and somewhat attempt to bridge the gap into the more general approach. The other speakers will carry the ball from there.

Basically the analysis of variance technique is just what the name implies - partitioning the variance (i.e. the square of the standard deviation) of an experiment into parts in order to test whether or not certain factors introduced into the design of the experiment actually produce significantly different results in the variable tested. That is, for example, does the mold in which a casting was made affect the porosity of the casting? Does the scale used affect the weight of a vial of medicine? In each case, we are interested in testing whether the effect of the factor on the variable measured is significant when compared with the random variation in the process. Hence the F test or variance ratio is used to make such comparisons.

Let us consider a specific example adapted from some unpublished data collected at Purdue. Here the variable measured (say X) is the rate of fluid flow in cubic centimeters and we are interested only in the effect of one factor, nozzle type, on this rate of flow. The results in Table I are for five runs through each of the 3 nozzle types.

TABLE I - Number of c.c. of Fuel thru 3 Nozzles for 5 trials.

		NOZZLE TYPE		
		A	B	C
Trial Number	1	96.6	96.6	97.0
	2	97.2	96.4	96.0
	3	96.4	97.0	95.0
	4	97.4	96.2	95.8
	5	97.8	96.8	97.0

The factor, nozzle type, is said to be in 3 categories as we have just 3 nozzles. It is assumed that these are the only nozzle types we are interested in. We do not wish to generalize our results to other nozzle types of which the 3 might be but a random sample. This is an important point as we are considering only these 3 nozzle types, we have a factor in fixed categories. Had we been interested in these 3 nozzle types as a random sample of a whole population of nozzle types, nozzle types would be a random effect. In a one-way classification (one factor) like this one, the analysis used to get the results would be the same for

either a random or a fixed effect but the significance tests performed would be interpreted differently. In this discussion, I shall confine my remarks to designs with fixed factors only, since Dr. Anderson will next be discussing a mixed model having both fixed and random factors in it.

For the one-way classification above, consider any observation - call it X_{ij}, where i is the number of the observation and j the category of the factor designation, i.e. the column in which X_{ij} is located. Thus $X_{23} = 96.0$ is the second observation for the third nozzle type. Now this number is affected by the category (or column) it is in, and also by some random error. We might then set up the model:

$$X_{ij} = \overline{X}' + B'_j + Z'_{ij}$$

where $\overline{X}'$ is a common term in all populations from which the experimental data were drawn, B'_j is a term designating the category or bias of the j'th factor from which the observation came (j = 1, 2, 3, in our problem) and Z'_{ij} is a random error. Our problem is to test the hypothesis that all 3 of these categories (or columns) came from like populations, or that: $B'_1 = B'_2 = B'_3 = 0$. If the least squares technique is applied to this model above, we find, for a sample of N observations:

$$X_{ij} = \overline{\overline{X}} + (\overline{X}_j - \overline{\overline{X}}) + (X_{ij} - \overline{X}_j)$$

where: $\overline{\overline{X}}$ is the best estimate of the population mean $\overline{X}'$.

$\overline{X}_j - \overline{\overline{X}}$ is the best estimate of the population B'_j.

and $X_{ij} - \overline{X}_j$ is the best estimate of the error or residual.

We note that this expression is a mathematical identity as both sides of the equation are equal for all values of X_{ij}.

Transposing this identity:

$$(X_{ij} - \overline{\overline{X}}) = (\overline{X}_j - \overline{\overline{X}}) + (X_{ij} - \overline{X}_j)$$

which says that the deviation of each observation (X_{ij}) from the grand mean of all observations in the experiment ($\overline{\overline{X}}$) can be broken down into two parts: the deviation of the category mean from the grand mean plus the deviation of observation from its own category mean.

Squaring both sides of this identity and summing first over all observations within a category and then over all k categories, the resulting expression is:

$$\sum_j\sum_i (X_{ij} - \overline{\overline{X}})^2 = \sum_j\sum_i (\overline{X}_j - \overline{X}..)^2 + 2\sum_j\sum_i (\overline{X}_j - \overline{X}..)(X_{ij} - \overline{X}_j)$$

$$+\sum_j\sum_i (X_{ij} - \overline{X}_j)^2$$

or, $$\sum_j\sum_i (X_{ij} - \overline{\overline{X}})^2 = \sum_j\sum_i (\overline{X}_j - \overline{\overline{X}})^2 + \sum_j\sum_i (X_{ij} - \overline{X}_j)^2$$

The middle term on the right hand side has vanished in the second expression since the summation over this cross product term equals zero. This expression says that the total sum of squares (of deviations) equals the sum of squares among category means plus the sum of squares within the categories. When each of these sums of squares (S.S.) is divided by the proper number of degrees of freedom, the quotient represents an unbiased estimate of the population variance from which the data came. If our hypothesis is true that the nozzle types do not affect the mean rate of flow then either the sums of squares among categories or the sums of squares within categories could be used to estimate the population variance, and each estimate is of the same variance and the ratio of the two estimates will follow an F distribution. Hence an F test using the proper number of degrees of freedom will test the hypothesis that the means of all 3 categories are from the same population or that $B_j = 0$ for all j.

Now the actual calculations are not made by substracting and getting the square of all these deviations but by making use of the binomial expansion of $\sum_i \sum_j (X_{ij} - \bar{\bar{X}})^2$. We can set up three rules for analyzing the effect of any factor on a given variable:

1) Square all the observations in the experiment and add, then subtract from this the square of the sum of all observations divided by N. This we call the total sum of squares. The sum of all observations squared and divided by N is often referred to as a correction term or "C.T." Associated with the total S.S. are N-1 degrees of freedom.

   [Symbolically this rule is: $\sum_i \sum_j X^2_{ij} - \frac{(\sum_i \sum_j X_{ij})^2}{N}$]

2) Sum all observations for each category of a given factor, square this total and divide by the number of observations for this category, sum for all categories, then subtract the correction term as in (1). This is called the S.S. for the given factor and has k-1 degrees of freedom associated with it, where k is the number of categories given.

   [Symbolically this rule is: $\sum_j \frac{(T_j)^2}{n_j} - \frac{(\sum_i \sum_j X_{ij})^2}{N}$]

   where T_j is the total of all n_j observations in category j.

3) Subtract the S.S. for the factor (or factors) from the total S.S. This is the residual or error S.S. The degrees of freedom are $(N-1)-(k-1) = N-k$ for the design with just one factor.

The data are then summarized in a table. Let us try this on our rate of flow problem. It will simplify the calculations considerably if the data are first coded. This can be done by subtracting 96.0 from each observation and then multiplying each observation by 10. The results are all integers as shown in Table 2:

TABLE 2 - Coded Data from Table 1.

		A	B	C	
Trial Number	1	6	6	10	
	2	12	4	0	
	3	4	10	-10	
	4	14	2	- 2	
	5	18	8	10	
Sums:		54	30	8	92
Sums of Squares:		716	220	304	1240

Total S.S. =

$$1240 - \frac{(92)^2}{15} = 675.73$$

$$\text{Nozzle S.S.} = \frac{(54)^2 + (30)^2 + (8)^2}{5} - \frac{(92)^2}{15} = 211.73$$

Source of Variation	Sum of Squares S.S.	d.f.	Mean Squares M.S.	F	$F_{.05}$
Total	675.73	14			
Among Nozzles	211.73	2	105.86	2.74	3.89
Error	464.00	12	38.67		

Applying the rules, the total S.S. is 675.73. For the effect of the 3 nozzle types, applying step (2) we add all readings for each nozzle type, square, divide by 5, and add for all 3 nozzles minus the correction term, giving 211.73. (Actually, here we divided by 5 after summing over all 3 types as the numbers of observations for each category are equal.) The remainder is 675.73 - 211.73 = 464.00. This is the error S.S. which is used as our 'yardstick' to test the significance of the different nozzle types on the data, i.e. to test the hypothesis that $B'_1 = B'_2 = B'_3 = 0$.

The data are summarized at the bottom of Table 2. The mean square is the unbiased estimate of the population variance found by dividing each S.S. by the associated number of degrees of freedom, and F is the ratio of the among nozzles mean square to the error mean square. In this problem F = 2.74. Consulting an F table with 2 and 12 d.f. the F necessary to claim a significant difference between nozzle types at the 5 °/o level of significance is $F_{0.05} = 3.89$. Since our F is less than this, we conclude that our hypothesis cannot be rejected, and we behave as though the 3 nozzle types produce no differences in the mean rate of flow.

Now some bright boy in the organization notes that five different operators worked these three nozzles and the data could be analyzed further for possible differences in rate of flow due to the different operators. The problem now becomes an analysis of variance with a

two-way classification of the data; i.e., two factors: nozzle type and operator, one in 3 categories (3 nozzle types) and the other in 5 categories (5 operators). Again we are assuming that the 5 operators are the only operators we are interested in, i.e. operators are a fixed factor. As each operator has worked with each nozzle type, we can analyze the data for differences in rate of flow among operators as well as among nozzle types. The model we assume now is:

$$X_{ij} = \bar{X}' + B'_j + C'_i + Z'_{ij}$$

where C'_i has been introduced to account for possible differences among operators where i = 1, 2, 3, 4, and 5. The analysis proceeds as before except for analysing the 5 categories of the operator factor. In Table 3, we have the analysis and the summary:

TABLE 3 - Two-way Classification: Nozzle Types and Operators

NOZZLE TYPE

		A	B	C	Sums
Operator Number	1	6	6	10	22
	2	12	4	0	16
	3	4	10	-10	4
	4	14	2	- 2	14
	5	18	8	10	36
	Sums	54	30	8	92

Operator S.S. =

$$\frac{(22)^2+(16)^2+(4)^2+(14)^2+(36)^2}{5} - \frac{(92)^2}{15} = 185.06$$

Source of Variation	S.S.	d.f.	M.S.	F	$F_{.05}$
Total	675.73	14			
Among Nozzles	211.73	2	105.86	3.03	4.46
Among Operators	185.06	4	46.26	1.33	3.84
Error	278.94	8	34.87		

Here we applied the same rule (No. 2 above) to each operator - add for each operator, square, divide by 3, and sum for all 5 operators minus the correction term.

$$\left[\text{Symbolically: } \sum_i \frac{(T_i)^2}{k} - \frac{(\sum_i \sum_j X_{ij})^2}{N}\right]$$

This source of variation or S.S. among operators is also subtracted from the total S.S. leaving a different and 'purer' error term than before. If we now recalculate the mean squares (that between nozzles stays the same), we can compare the among nozzles mean square with our revised estimate of the error mean square and also can compare the operator mean square with this error mean square. The F values are now as given in Table 3 along with their 5 °/o significance F values. The results show that neither the nozzles nor the operators produce a significant difference in the average rates of flow even though we have now reduced the error term by accounting for another possible source of variation. In the first analysis (one-way classification), the operator effects were included in (that is, "confounded") with the error term. In actual

practice, this other source of variation should have been foreseen in the original design and the second model used as shown in Table 3. We thus fail to reject the hypothesis that: $B'_1 = B'_2 = B'_3 = 0$, and also fail to reject the hypothesis that $C'_1 = C'_2 = C'_3 = C'_4 = C'_5 = 0$.

A further extension of our analysis would be possible if we were to repeat the same experiment, thus getting at least two observations for each operator-nozzle combination. This we call a replication and it is assumed that both (or all) replications are taken under the same conditions. Replication will enable us to analyze for a possible interaction between the two main factors, i.e. between nozzle types and operators. The replication would yield a still better estimate of the error in the experiment than was possible before. Of course this also requires more data. The model would now be:

$$X_{ij} = \overline{X}' + B'_j + C'_i + D'_{ij} + Z'_{ij}$$

where D'_{ij} represents the possible interaction between certain categories of one factor with categories of the other factor. Repeating the above experiment three times the results (coded) might be summarized as in Table 4:

TABLE 4: Nozzles vs. Operators Data with Replication

Operator

1			2			3			4			5		
Nozzle			Nozzle			Nozzle			Nozzle			Nozzle		
A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
6 6 -15	13 6 13	10 10 -11	26 12 5	4 4 11	-35 0 -14	11 4 4	17 10 17	11 -10 -17	21 14 7	-5 2 -5	12 - 2 -16	25 18 25	15 8 1	-4 10 24

You will notice that this table has been set up somewhat differently than the usual 3 by 5 table as Table 3. It is hoped that this will give a clearer idea of what interaction is and how its S.S. is computed. If we forget about nozzle types and operators for the moment and analyze the above as a one-way classification of 15 cells we find, applying the same rules set down earlier that:

The total S.S. = 7085.24
Among cells S.S. = 4047.24
Within cells S.S. = 3038.00

Now, the among cells variation includes variation among operators and among nozzle types as well as chance variation. If we calculate the S.S. for each of these main effects, we find:

Among nozzles S.S. = 1426.97
Among operators S.S. = 798.79

When these two main effect S.S. are substracted from the among cells S.S., we find that there is some variation left over (1821.48),

which is unaccounted for by the nozzle types or by the operators. This is what we label the nozzle-operator interaction or NxO interaction S.S. This is a first order interaction as it is the simplest type observable in any experiment. It represents variation between the means of these "cells" not attributable to either of the main effects and is present because of the way certain operators might interact with certain nozzle types. For example, the second operator might run nozzle type A consistently too fast while running type C consistently too slowly and this would not show up as a nozzle or operator effect if another operator tended to reverse this by running nozzle C too slowly and A too fast. We now summarize for Table 4:

TABLE 5: Summary of Nozzle, Operator Data with Replication

Source of Variation	Sum of Squares			d.f.	Mean Square	F	$F_{.05}$
Total	7085.24			44			
Among cells	4047.24	Among Nozzles	1426.97	2	713.49	7.05	3.32
		Among Operators	798.79	4	199.70	1.97	2.69
		NXO Interaction	1821.48	8	227.68	2.25	2.27
Within cells (error)	3038.00		3038.00	30	101.27		

Whether to pool or not to pool the error and interaction for testing the main factors is still debatable among the experts, but we can test the significance of the interaction mean square with the error mean square as the 'yardstick'. If this is not significant, we can test the two main factors versus this error term or possibly pool the interaction and error terms as the yardstick for testing these main factors. The results above show the nozzle types producing a highly significant difference in mean rate of flow while operators show no significant differences, nor is the interaction quite significant if we use the 5 o/o significance level.

In this discussion, we have tried to set up some general rules for getting the sums of squares and to review the analysis of variance technique. These general methods can easily be extended to analyze as many factors as you wish to assume in your original model, and also to compute the interaction of the first order or higher. It should be noted, however, that second order interactions are what is left in the among cells S.S. (among cubes, really) after we subtract out the S.S. for the 3 main factors <u>and</u> all 3 first order interactions. The designs illustrated here are all factorial designs where each category of one factor is combined with each category of every other factor.

Finally, some mention should be made of the assumptions underlying the analysis of variance technique. The basic assumptions are:

1) The effect of all factors is additive. We have used linear models throughout this discussion.

2) The random errors were sampled from a normal universe.

Often some transformation of the variable can be made if this assumption

is not met. How often we see a discrete variable used for analysis such as the number of defective items produced by several machines from several batches of raw material. Such data are likely to be from a binomial distribution, and the normality assumption is only approximately met if the average number of defectives is quite high. However, by using the percent defective as the variable and transforming it by an arc sine transformation, the normality assumption may be better satisfied.

3) The data must exhibit homoscedasticity or homogeneity of variance. That is, we should first show that no significant differences exist among the variances within the cells. This is usually tested with the Bartlett test. When transformations are made to induce normality, we may also expect more homogeneity of variance.

QUALITY INCENTIVE

John F. Wagner
College of Engineering, University of Colorado

We have had, for some time, quantity incentive plans in manufacturing plants, in business office routines, and in other places as well.

In some of these applications, at least, it is felt that quality is sacrificed to the economic pressure for quantity. Many operators feel, perhaps not without reason, that their overall return is maximized by increasing their rate of production even though this results in their having to repair or to replace the augmented number of defectives "on their own time." "On their own time" means that they cannot produce additional units while they repair or replace the previously produced defective ones.

The tendency to produce quantity at the expense of quality is enhanced if the unit of the operator's work is accepted or rejected on the basis of a sampling scheme.

To combat this inclination on the part of the operator it is felt, in some applications at least, that a quality incentive would be helpful. A general formula would be most useful if one could be written, but a specific example and the bases for its details may be of interest and help.

BACKGROUND OF THE PLAN

In the embossing operation of the Esquire-Coronet Subscription Fulfillment process, the subscriber's name, the subscriber's address, and a subscription code are impressed on a Speedaumat plate for servicing of the subscriber's order.

An error rate of one per cent has been set as the standard for this operation. This is interpreted to mean that an operator suited to this type of work should, after proper training and a reasonable time for experience, attain and maintain an overall error rate of not more than one defectively embossed plate out of 100 produced.

The unit of the operator's work is accepted or detailed on the basis of the Dodge-Romig SA-1.0 table. This guarantees, if the proof-reading is properly done, that the long run proportion of defective plates in the files will not be greater than one per cent, regardless of the individual operator's error rate.

The bases for quality incentive come, in this application, from two considerations. Because of the present labor situation, the company feels that it must be content with an embossing error rate greater than the standard set -- a value of, say, two per cent. If the operator's rate of error production is less than this reference value, the cost of proof-reading is thereby reduced. This reduction is a result of the lessening of the frequency of detailing which accompanies a lowered error rate. The decrease in detailing lowers the ratio of the average number of plates inspected to the total number produced. This ratio approaches a limiting value (which depends upon the lot size) as the error rate goes to zero.

The other basis for such a quality incentive has to do with complaints. The subscriber may occasionally have reason to complain about non-delivery of magazines, incorrect termination of his subscription, failure to send a card to the recipient of his gift, et cetera. Some of these "squawks" arise from errors in embossing which get through the proof-reading operation as a result of an AOQL acceptance plan (and perhaps, too, due to errors in proof-reading). But fewer errors will be passed on to the subscriber if the operator does his work more accurately, and the subsequent lowering in number of complaints not only represents a saving to the company in adjusting effort, but also in subscriber good will.

We use this set of symbols:

p = operator's error rate in percent defective (determined on the basis of SA-1.0 samples only)

$\bar{p}$ = expected error rate (2%) an arbitrary figure based upon observation of what is happening, or upon that operator error performance which will be regarded as satisfactory

n = sample size (determined from SA-1.0 table)

N = lot size (usually 150 plates in this application)

P_a = probability of accepting by a sampling plan, a lot of size N which is p-percent defective.

I = average number of plates inspected per lot of size N and of quality p

AOQ = average outgoing quality (going out from the inspection operation, provided correct proof-reading was done) in percent defective

AOQL = average outgoing quality limit

c = acceptance number

i = average cost in cents of proof-reading each plate

a = average cost in dollars of adjusting each complaint

L = number of lots produced in a given time (say, one week)

r = ratio of complaints to errors existing on plates in files

Preliminary computations are illustrated in Table I. This table shows the I values and the AOQ values corresponding to representative values of p in connection with a sampling acceptance/rectification plan-Dodge and Romig's SA-1.0 plan, 1% AOQL, with $N = 150$, $n = 32$, $c = 0$. P_a values were read from Table G, in E. L. Grant's <u>Statistical Quality Control</u> (McGraw-Hill). This assumes that the Poisson Exponential Binominal Limit is a sufficiently good approximation to these probabilities. Rounding off was done upward, even for values somewhat below the usual separation line, as a proof reader is more likely to "miss" a defective than to call a good plate bad.

TABLE 1 - Computation of I/N and AOQ for Representative Values of p
AOQL = 1%; N = 150; n = 32; C = 0

p (%)	np	P_a	I	I/N	AOQ (%)
0.0		1.000	32.00	0.2133	0
0.1	0.032	0.969	35.66	0.2377	0.0762
0.2	0.064	0.938	39.32	0.2621	0.1476
0.3	0.096	0.909 (0.908)	42.74	0.2849	0.2145
0.4	0.128	0.881 (0.87)	46.04	0.3069	0.2772
0.5	0.160	0.853 (0.85)	49.35	0.3290	0.3355
0.6	0.192	0.826 (0.83)	52.53	0.3502	0.3899
0.7	0.224	0.800 (0.80)	55.60	0.3707	0.4406
0.8	0.256	0.775 (0.77)	58.55	0.3903	0.4878
0.9	0.288	0.750 (0.75)	61.50	0.4100	0.5310
1.0	0.320	0.727 (0.73)	64.21	0.4281	0.5719
1.1	0.352	0.704 (0.70)	66.93	0.4462	0.6092
1.2	0.384	0.681 (0.68)	69.64	0.4643	0.6430
1.3	0.416	0.660 (0.66)	72.12	0.4808	0.6750
1.4	0.448	0.640 (0.64)	74.48	0.4965	0.7049
1.5	0.480	0.620 (0.62)	76.84	0.5123	0.7317
1.6	0.512	0.600 (0.60)	79.20	0.5280	0.7552
1.7	0.544	0.581 (0.58)	81.44	0.5429	0.7771
1.8	0.576	0.563 (0.56)	83.57	0.5571	0.7972
1.9	0.608	0.545 (0.54)	85.69	0.5713	0.8147
2.0	0.640	0.528 (0.53)	87.70	0.5846	0.8308
2.5	0.800	0.449 (0.44)	97.02	0.6468	0.8833
3.0	0.960	0.383 (0.38)	104.81	0.6987	0.9039
3.5	1.120	0.327 (0.33)	111.41	0.7428	0.9006
4.0	1.280	0.279 (0.27)	117.08	0.7805	0.8780
4.5	1.440	0.238 (0.24)	121.92	0.8128	0.8429
5.0	1.600	0.202 (0.20)	126.16	0.8411	0.7945

It is interesting to note how close to these P_a values one comes by use of a chart such as the one in Dodge and Romig's book for the cumulative probability values—Poisson Exponential. The values given under the P_a column in the parentheses, are those obtained from this curve with just the eye as a guide, no ruler or other scaling device used. The curve-obtained quantities were written down a day or two before Table G values were read off. However, those read from the chart give no more than two-digit accuracy in P_a and in I.

I is computed from the relation:

$$I = nP_a + (1 - P_a)\,N \qquad (1)$$

and AOQ from the formula:

$$AOQ = p\,(1 - I/N) \qquad (2)$$

the defectives being replaced by good plates. I/N is included not only because it makes the following formulas somewhat more symmetrical, but principally because the I/N value does not change much as N varies within wide limits (from N = 150 to N = 250, I/N goes from 0.58 to 0.54 for p = 2.0%).

Although four digits have been retained for the recorded values for I/N and for AOQ, it should be borne in mind that their accuracy is limited by the interpolation for P_a, and subsequent operations. I is accurate to within ± 0.001. AOQ values have varying absolute accuracies, from $\pm$ 0.0006 at p = 0.7%, to $\pm$ 0.0016 at p = 2.0%, to $\pm$ 0.0032 at p = 4.0%. These values are based on the assumption that the p-values are known exactly and that the interpolation for P_a is subject to an error of as much as $\pm$ 0.001-$\pm$ 0.0005 from rounding off at time of construction of the table, and $\pm$ 0.0005 from rounding off at interpolation.

We write down, now, the two formulas for savings realized by the company due to improvement of quality performance with respect to $\bar{p}$. Both of these expressions are for a standard length of time, say, one week.

Saving (cents) due to decreased ($p < \bar{p}$) inspection
$$= \left[(I/N)_{\bar{p}} - (I/N)_{p}\right] Nil \qquad (3)$$

Saving (cents) due to fewer ($p < \bar{p}$) complaints
$$= \left[(AOQ)_{\bar{p}} - (AOQ)_{p}\right] NaLr \qquad (4)$$

Out of all of this, the following plan seemed to have some merit and has been proposed to the Subscription Fulfillment Division of Esquire-Coronet. It is only fair to acknowledge that many of the details of this plan are a result of conferences with the subscription manager, Mr. J. L. Ross, and the industrial engineer, Mr. P. C. Miller, at Esquire-Coronet.

THE PROPOSED PLAN

Assuming that a certain portion (say 50 percent--some people don't know, for example, when their subscription should end and thus they renew upon notice from the publisher) of such errors as we count lead to complaints, we propose that a percentage of the savings realized by the company be turned over to the employee-embosser. For the purposes of this paper, let us say, 60 percent. Some of the difference could well

be used to upgrade the proof-reader, as his (her) work is of paramount importance to any plan based on a sampling scheme. The pay which an embosser receives would then be her base pay, a quantity bonus, and a quality bonus. She might receive a quality increment even though her quantity production is somewhat below norm (80 percent of standard production), and also, she may lose some, or all, of her quantity bonus if the quality for that week is bad (more than two percent). Thus the plan would establish a double deterrent to producing quantity at the expense of quality.

The current quantity incentive plan is given by the graph in Fig. 1. Pay Rate gives the percentage of base pay received (never less than 100 percent) and Operating Rate gives the percentage of standard production which the operator has attained for the week.

As used at present the Pay Percent (M) is given in terms of the Operating Percent (W) by the following functional relations:

$M = 100\%$ (base pay); $W \leq 80\%$*
$M = 0.5W + 60$; $80\% \leq W \leq 120\%$
$M = W$; $120\% \leq W$

*The percentages here are based on Standard Production = 100%.

For the proposed plan, it is suggested that this quantity incentive be modified as follows to allow for an overall bonus to an operator who performs accurately, even though somewhat below norm (80% of Standard) quantity-wise:

$M = 0.5W + 60$; $0\% \leq W \leq 120\%$
$M = W$; $120\% \leq W$

This is used with the understanding that the combined quality and quantity adjustment shall not be negative; that the gross pay for the week shall never be less than the employee's base rate.

Table II gives the necessary numerical details for operating the proposed plan on the basis of the following assumptions:

1. The operator's unit of work is 150 plates

2. The standard rate of production (100 percent operating performance) is 71.5 plates per hour

3. The operator's base pay is $40.00 per week

4. The cost of inspection is one cent per plate, based on direct labor and supervision.

5. The cost of adjusting a complaint or request is $0.65, same basis as in (4)

6. The basic (reference) error rate is taken as two percent and all quantity bonus is wiped out if error rate for the given week is found to be four percent or over as shown by the samples taken on the basis of the Dodge-Romig SA-1.0 plan.

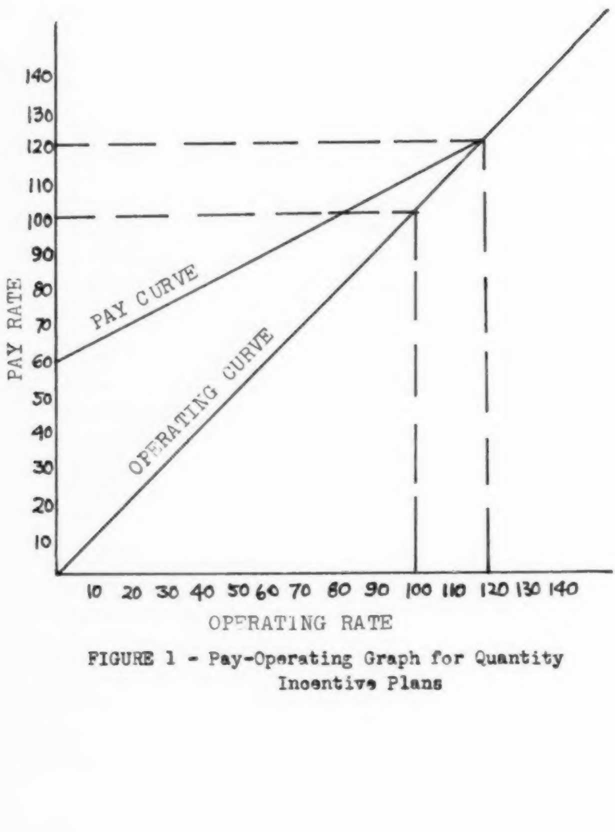


FIGURE 1 - Pay-Operating Graph for Quantity Incentive Plans

7. The percentage of deduction (D) is a quadratic function of the error rating; for $2 \leq p \leq 4$, $D = 20p^2 - 70p + 60$ such that $D = 0$ for $p = 2$ and $D = 100\%$ for $p = 4$.

The table shows contributions and total bonus (in dollars) for quantity (Formula 5) and quality (Formulas 3 and 4) considerations based on the above assumptions.

Calculations have been made for typical percentages of Operating Standard (70, 80, 100, 120, 130) and selected values for Operating Error Rate, p.

TABLE II - Dollar Increments to Base Wage of $40.00 per Week Due to Quantity and Quality Sources at Selected Percentages of Quantity Production and for Selected Levels of Quality

Operator Error Rate	Percent of Operating Standard											
	70				80				100			
	Quan.	QUALITY		Bonus	Quan.	QUALITY		Bonus	Quan.	QUALITY		Bonus
p (%)	(5)	(3)	(4)	Total	(5)	(3)	(4)	Total	(5)	(3)	(4)	Total
0.0	-2.00	4.46	3.24	5.70	0.00	5.10	3.70	8.80	4.00	6.37	4.63	15.00
0.1	-2.00	4.17	2.95	5.12	0.00	4.76	3.37	8.13	4.00	5.95	4.21	14.16
0.2	-2.00	3.87	2.67	4.54	0.00	4.43	3.05	7.48	4.00	5.53	3.81	13.34
0.4	-2.00	3.34	2.16	3.50	0.00	3.81	2.47	6.28	4.00	4.77	3.09	11.86
0.6	-2.00	2.82	1.72	2.54	0.00	3.22	1.97	5.19	4.00	4.02	2.46	10.48
0.8	-2.00	2.33	1.34	1.67	0.00	2.67	1.53	4.20	4.00	3.33	1.91	9.24
0.9	-2.00	2.10	1.17	1.27	0.00	2.40	1.34	3.74	4.00	3.00	1.67	8.67
1.0	-2.00	1.88	1.01	0.89	0.00	2.15	1.16	3.31	4.00	2.69	1.44	8.13
1.1	-2.00	1.66	0.87	0.53	0.00	1.90	0.99	2.89	4.00	2.37	1.24	7.61
1.2	-2.00	1.45	0.73	0.18	0.00	1.65	0.84	2.49	4.00	2.06	1.05	7.11
1.4	-2.00	1.06	0.49	-	0.00	1.21	0.56	1.77	4.00	1.51	0.70	6.21
1.6	-2.00	0.68	0.30	-	0.00	0.78	0.34	1.12	4.00	0.97	0.42	5.39
1.8	-2.00	0.33	0.13	-	0.00	0.38	0.15	0.53	4.00	0.47	0.19	4.66
2.0	-2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.00	0.00	0.00	4.00
2.5	-2.00	-	-	-	0.00	-	-	-	4.00	Less	10%	3.60
3.0	-2.00	-	-	-	0.00	-	-	-	4.00	Less	30%	2.80
3.5	-2.00	-	-	-	0.00	-	-	-	4.00	Less	60%	1.60
4.0	-2.00	-	-	-	0.00	-	-	-	4.00	Less	100%	0.00

TABLE II (Continued)

Operator Error Rate p (%)	Percent of Operating Standard							
	120				130			
	Quan.	QUALITY		Bonus	Quan.	QUALITY		Bonus
	(5)	(3)	(4)	Total	(5)	(3)	(4)	Total
0.0	8.00	7.65	5.56	21.21	12.00	8.28	6.02	26.30
0.1	8.00	7.14	5.05	20.19	12.00	7.74	5.47	25.21
0.2	8.00	6.64	4.57	19.21	12.00	7.19	4.95	24.14
0.4	8.00	5.72	3.70	17.42	12.00	6.19	4.01	22.20
0.6	8.00	4.83	2.95	15.78	12.00	5.23	3.20	20.43
0.8	8.00	4.00	2.30	14.30	12.00	4.33	2.49	18.82
0.9	8.00	3.60	2.01	13.61	12.00	3.89	2.17	18.06
1.0	8.00	3.22	1.73	12.95	12.00	3.49	1.88	17.37
1.1	8.00	2.85	1.48	12.33	12.00	3.09	1.61	16.70
1.2	8.00	2.48	1.26	11.74	12.00	2.68	1.36	16.04
1.4	8.00	1.81	0.84	10.65	12.00	1.96	0.91	14.87
1.6	8.00	1.17	0.51	9.68	12.00	1.26	0.55	13.81
1.8	8.00	0.57	0.22	8.79	12.00	0.61	0.24	12.85
2.0	8.00	0.00	0.00	8.00	12.00	0.00	0.00	12.00
2.5	8.00	Less	10%	7.20	12.00	Less	10%	10.80
3.0	8.00	Less	30%	5.60	12.00	Less	30%	8.40
3.5	8.00	Less	60%	3.20	12.00	Less	60%	4.80
4.0	8.00	Less	100%	0.00	12.00	Less	100%	0.00

EXPERIMENTS WITH MANY FACTORS

K. A. Brownlee[1]
University of Chicago

Simple Experiment

For a long time the approved method of investigating a process with a number of variables was to vary them one at a time. If consideration is given to a process with six variables, and the testing of these only at two levels each, then the two possible levels of each factor can be represented by the symbols 0 and 1. Thus, the symbol 000000 denotes that set of conditions with all six factors at their lower level, and the symbol 010100 represents that set with the second and fourth factors at their upper levels and the remaining factors at their lower levels. If one were to investigate the effects of these six factors according to the traditional principles, then observations would be made with the following sets of conditions:

000000, 100000, 010000, 001000, 000100, 000010, 000001 (1)

Here the first symbol represents the so-called "control" or standard, and the succeeding six represent trials with each variable altered in turn, the remaining five variables being held constant at their lower, or standard, levels. On the basis of a comparison of each experimental trial with the control, a decision on the use of that factor at its upper level or its lower level would be made. The over-all decision would be the sum of the individual decisions.

Factorial Experiment

Sometimes the foregoing procedure would not prove satisfactory. A study of data in Table 1 is suggested. Figures therein represent mean cross-sectional areas for cakes of two types baked from batters with two pH's. If a simple experiment had been executed, there would be only those observations marked with an asterisk. The effect of changing from low to high acidity keeping the type of cake at chocolate, gives a larger cake area; and the effect of changing the kind of cake, keeping the acidity at its low value, also gives a larger cake area. Therefore, both these factors at their upper levels should be used, but actually this condition is less optimal than the high acidity, chocolate-cake combination. An effect of this type, where the separate effects of the two factors are not additive, is known to the statistician as an interaction. Another way of looking at it is to observe that the effect of one factor varies according to the level of the other factor. Interactions involving two factors are known as first-order interactions. If a first-order interaction between two factors A, and B, varies according to the level of a third factor C, then this is called a second-order interaction and denoted symbolically as A x B x C. The relationship between three factors involved is symmetrical, so that the interaction between A and C, varying according to the level of B, and the interaction between B and C, varying according to the level of A, are identical with the A x B x C interaction.

---

[1]This work was sponsored by the Army, Navy and Air Force through the Joint Services Advisory Committee for Research Groups in Applied Mathematics and Statistics by Contract No. N6ori-02035. Reprinted with permission from Chemical Engineering Progress, Vol. 49, pp. 617-621, 1953.

Table 1

	Chocolate (C_0)	White (C_1)	C_1-C_0
Low acidity (A_0)	36.87*	42.44*	5.56
High acidity (A_1)	51.56*	48.44	-3.12
$A_1 - A_0$	14.69	6.00	

In order to detect and measure the presence of interactions in a multifactor system, one must make, not the simple one-variable-at-a-time experiment, but a factorial experiment which consists of all combinations of all factors. In addition to being able to detect interactions, it has the advantage, where interactions are absent, of greater efficiency than the simple experiment. This efficiency arises through the ability to take averages over all observations for estimating the effects of each of the factors. This is in contrast to the simple experiment where each observation, other than the control, serves to provide information only on the factor that was varied for that particular trial (7).

Algebraic Representation of Effects

To consider how the factorial experiment uses all its observations to estimate the average effect of a factor, the following symbolism is introduced: The small letters a, b, c, etc., denote that the factor in question is at its upper level, and the absence of a letter implies that the factor is at its lower level. Thus, in a six-factor experiment the treatment combination a c e has factors A, C, and E at their lower levels and factors B, D, and F at their lower levels. The symbol (1) denotes the treatment combination with all factors at their lower levels. In the previous notation a c e would represent 101010. A slight ambiguity in notation can be tolerated: symbols of the type a c e will be used to represent both the particular combination of levels and also the numerical result, be it yield or purity or some other property, obtained from that experimental condition.

For simplicity, a factorial experiment with three factors is considered. There will be $2^3 = 8$ treatment combinations, namely, (1), a, b, ab, c, ac, bc, abc. There will be four estimates of the effect of A, as in Table 2. The average effect of these four estimates is defined as the main effect of A,

$$\begin{aligned} A &= \tfrac{1}{4}\left[(a-1) + (ab-b) + (ac-c) + (abc-bc)\right] \\ &= \tfrac{1}{4}(a-1)(b+1)(c+1). \qquad (2) \end{aligned}$$

There will be two possible first-order interactions involving A, namely, AB and AC. To consider AB, it is defined as one-half the mean difference between the effect of A with B at its upper level (averaged over both levels of c),

$$\tfrac{1}{2}\left[(abc-bc) + (ab-b)\right]$$

and of A with B at its lower level (averaged over both levels of C),

$$\tfrac{1}{2}\left[(ac-c) + (a-1)\right].$$

Table 2

Level of B	Level of C	Estimate of a
Lower	Lower	a - (1)
Upper	Lower	ab - b
Lower	Upper	ac - c
Upper	Upper	abc - bc

This difference, divided by two, is

$$AB = \tfrac{1}{2}\left[\tfrac{1}{2}\{(abc-bc) + (ab-b)\} - \tfrac{1}{2}\{(ac-c) + (a-1)\}\right]$$

$$= \tfrac{1}{4}(a-1)(b-1)(c+1) \qquad (3)$$

One could have considered AB as one-half the mean difference between B with A at its upper level (averaged over both levels of C) and B with A at its lower level (averaged over both levels of C). This would have led to the same result (3).

A second-order interaction involving all three factors is conceivable. One way to obtain it is to consider one-half the mean difference between the AB interaction for the upper level of c,

$$\tfrac{1}{2}\left[(abc-bc) - (ac-c)\right]$$

and the same for the lower level of c,

$$\tfrac{1}{2}\left[(ab-b) - (a-1)\right] .$$

Half the difference between these two is

$$ABC = \tfrac{1}{2}\left[\tfrac{1}{2}\{(abc-bc) - (ac-c)\} - \tfrac{1}{2}\{(ab-b) - (a-1)\}\right]$$

$$= \tfrac{1}{4}(a-1)(b-1)(c-1). \qquad (4)$$

The alternative viewpoints will lead to the same result.

The general method of writing down the arrangement of treatment combinations to give an estimate of any specified effect will be obvious from an inspection of (2), (3), and (4). All letters occur on the right-hand side of these equations in the form $(a\pm1)(b\pm1)(c\pm1)$. If a letter occurs on the left-hand side, i.e., if α, β, or $\gamma = 1$, then the corresponding sign in the bracket on the right-hand side is minus, otherwise plus. In the general case where there are n factors, the divisor is 2^{n-1}, and one has

$$A^{\alpha}B^{\beta}C^{\gamma} \cdots = \frac{1}{2^{n-1}} \left[a-(-1)^{\alpha+1}\right] \left[b-(-1)^{\beta+1}\right] \left[c-(-1)^{\gamma+1}\right] \cdots \qquad (5)$$

It will be noted that the expressions (2), (3), and (4) for the various effects each employs all results of the experiment. Thus, in the case of a six-factor experiment, if the interactions are not significant, the main effect of A is estimated as the difference between two averages each of thirty-two observations. If the simple experiment

represented by expression (1) had been replicated nine times, it would have required sixty-three runs, almost exactly the same as the factorial experiment which required sixty-four, but its estimates of the factors are the differences between two averages each of nine observations. The ratio of the variance of the two experiments will then be 9/32 = 0.281, or if adjusted by the ratio of numbers of runs (64/63) x 0.281 = 0.296. The simple experiment thus has an efficiency in measuring main effects relative to the factorial of less than 30%. The larger the experiment, the greater the relative efficiency of the factorial arrangement.

Fractional Replication of Factorial Experiments

The main disadvantage of the factorial experiment is that the number of runs required becomes rather large for a large number of factors. Even when all factors are at two levels, a six-factor experiment requires sixty-four runs, an eight-factor experiment 256 runs. Such large numbers of runs are often greater than is either practical or necessary from the point of view of estimating the effects with a specified accuracy. The need exists, therefore, for designs which will retain the efficiency aspect of the factorial experiment and the capability of detecting the presence of interactions. The solution to this problem lies in replicating only a certain fraction of the whole factorial experiment. The technique is due to Finney (6) and the rationale is along the following lines. For fuller accounts see (2, 5, 9).

Next under consideration is a four-factor experiment in which only half the treatment combinations were used. The sets are those with a plus sign in the expression for ABCD. These will be

$$(1),\ ab,\ ac,\ bc,\ ad,\ bd,\ cd,\ abcd \qquad (6)$$

An estimate of the main effect of A, for example, will be the difference between the average of those with A at its upper level and the average of those with A at its lower level:

$$A = \tfrac{1}{4}\left[ab + ac + ad + abcd - (1) - bc - bd - cd\right] \qquad (7)$$

Examining the arrangement of results for estimating BCD, which would in a full replicate be

$$\begin{aligned}(a+1)(b-1)(c-1)(d-1) = \tfrac{1}{8}\ [&ab + ac + ad - bc - bd - cd \\ &-a + b + c + d - abc - abd - acd \\ &+ bcd + abcd - (1)] \qquad (8)\end{aligned}$$

However, in this half-replicated experiment there are only the eight treatment combinations given in (6). Using these, our estimate of BCD is

$$BCD = \tfrac{1}{4}\left[ab + ac + ad + abcd - (1) - bc - bd - cd\right] \qquad (9)$$

but this is identical with our estimate of A in (7).

Similarly if one expands $AB = (1/8)\ (a-1)(b-1)(c+1)(d+1)$ and $CD = (1/8)(a+1)(b+1)(c-1)(d-1)$ and selects the eight treatment combinations in (6), one arrives at the following:

$$AB = \tfrac{1}{4}\ \left[abcd + ab - ac - ad - bc - bd + cd + (1)\right]\ = CD \qquad (10)$$

If the $2^4 - 1 = 15$ effects in this way are written out, it will be found that they fall into seven pairs, making fourteen, with the fifteenth being that used for selecting the 2^3 treatment combinations, here ABCD. The pairings are

$$
\begin{array}{ll}
A = BCD & AB = CD \\
B = ACD & AC = BD \\
C = ABD & AD = BC \\
D = ABC &
\end{array}
\qquad (11)
$$

It can be seen that the rule for finding the alias of any effect is to multiply it by ABCD, using the rule that $A^2 = B^2 = C^2 = D^2 = 1$. For example,

$$A = A \times ABCD = A^2BCD = BCD.$$

From the practical point of view the half-replicate experiment just outlined is not satisfactory as it leads to appreciable ambiguity. The main effects with second-order interactions as aliases are probably all right, as if, for example, one found $A = BCD$ to be appreciable it would be much more probable to be A than BCD, as it is an empirical observation that main effects are more often appreciable than high-order interactions. The first-order interactions are, however, hopelessly confused.

As the number of factors in the experiment is increased, the risk of ambiguity decreases. When one gets to the six-factor experiment, which as a half-replicate requires thirty-two treatment combinations, it is an experiment adequate for many purposes. The aliases are of the type

$$
\begin{array}{rll}
A = & BCDEF & \text{(six such)} \\
AB = & CDEF & \text{(fifteen such)} \\
ABC = & DEF & \text{(ten such).}
\end{array}
$$

The second-order interactions would be used as estimates of error, and the danger of ambiguity in the main effects and first-order interactions is small. An example of this design applied to plant-scale penicillin fermentation has been given by Brownlee (3).

The great disadvantage of testing a factor at two levels only is that the existence or position of a maximum or minimum cannot be estimated. There are some factorial replicates which include one or more factors at four levels, the remainder being at two levels. The smallest experiment of this type which can be considered reasonably secure is the $4^1 \times 2^4$, i.e., one factor at four levels and four factors at two levels. This design is derived from the 2^6 by allocating two of the two-level factors to represent the four-level factor. The two pseudofactors are arranged as in Table 3. Thus $(1) = 1$, $a = 2$, $b = 0$, $ab = 3$, where 0, 1, 2, and 3 are levels of the four-level factor.

Table 3

	b_0	b_1
a_0	1	0
a_1	2	3

Half-replicate Example

As an example of this design the results of an experiment on the baking of a cake will be considered (10). The dependent variable under study was the average cross-sectional area in square inches. The cakes were of two kinds, chocolate and white (denoted by C_0 and C_1). They were made with one of two baking powders (sulfate-phosphate and tartrate, B_0 and B_1). The original experiment had three types of shortening, but only two are included here, butter and vegetable shortening, denoted by S_0 and S_1. The pH's of the mixtures were adjusted to either of two values and these are denoted by A_0 and A_1. There were four storage times. The experiment is thus $4^1 \times 2^4$, and the author fully replicated it. However, it can be analyzed as though it had been a half-replicate, in which case only the results marked with an asterisk in Table 4 would be available.

Table 4

		B_0				B_1			
		S_0		S_1		S_0		S_1	
		A_0	A_1	A_0	A_1	A_0	A_1	A_0	A_1
C_0	T_0	43	58*	39*	54	49*	67	50	59*
	T_1	38*	53	38	55*	44	58*	36*	51
	T_2	34	50*	34*	46	42*	56	31	45*
	T_3	30*	47	25	41*	33	49*	24*	36
C_1	T_0	45*	51	40	47*	49	57*	45*	53
	T_1	41	46*	42*	46	45*	54	47	54*
	T_2	41*	45	42	47*	45	50*	44*	49
	T_3	37	44*	37*	44	41*	43	38	45*

*Results available.

In a half-replicate one way of determining which is the appropriate set of treatment combinations is to use the method by which these designs were examined, namely, to expand the highest-order interaction and use only those treatment combinations with a plus sign. Here would be expanded:

$$(a-1)(b-1)(c-1)(d-1)(e-1)(f-1). \qquad (12)$$

The combinations of a and b would then have to be converted into levels of the four-level factor by using Table 3. Allocating T to a and b, B to c, S to d, A to e, and C to f, then the entry $T_0B_0S_0C_0T_0$ in Table 3 is a, which has a negative sign in the expansion of (12) and so is not included in the half-replicate. However, $T_3B_1S_1A_1C_1$ is abcdef, and this has a plus sign in the expansion of (12) and so is included. Although this method of finding the treatment combinations is the simplest, there is an alternative which is more expeditious (2, 6, 9).

The interpretation of the results of a complex experiment clearly requires some special techniques for just by looking at the data it is

difficult to determine what there is to see. The tool, the statistician uses, is the analysis of variance. Broadly speaking, this analyzes the variance, defined as the sum of the squares of the deviations of the observations from the mean, divided by the degrees of freedom, into its constituent components attributable to the various effects. Those components which are significantly larger than the error, are looked at; the remainder, in general, are forgotten. For the analysis of variance see (8, 13). The analysis of variance of data of Table 4 is in Table 5, both for the half-replicate and for the full replicate.

Table 5

Source of Variance		Sums of Squares	
		Half-Replicate	Full Replicate
Chocolate vs. White	(C)	16.531	23.767
Baking Powder	(B)	75.031	185.641
Shortening	(S)	75.031	159.391
pH	(A)	935.281	1711.891
Storage time	(T): L	507.656	1212.903
	Q	9.031	13.141
	C	3.906	8.128
C x B		7.031	5.641
C x S		34.031	83.266
C x A		166.531	301.891
B x S		26.281	37.516
B x A		2.531	0.391
S x A		3.781	1.891
C x T:	L	79.806	196.878
	Q	0.031	3.516
	C	0.006	0.153
B x T:	L	12.656	71.253
	Q	1.531	0.141
	C	0.156	2.278
S x T:	L	0.306	1.378
	Q	16.531	15.016
	C	10.506	5.778
A x T:	L	0.506	0.903
	Q	1.531	1.266
	C	28.056	3.003
Remainder		20.439*	184.093**
Total		2034.719	4231.109
Residual Mean Square		3.4065*	4.8446**

* With 6 degrees of freedom.
** With 38 degrees of freedom.

In the present instance the four-level factor represents a quantitative variable, being storage time. If the assumption is made that the intervals between successive levels are equal, then it becomes easy to test the fit of a polynomial equation of the form

$$y = a + bx + cx^2 + dx^3.$$

In the analysis of variance in Table 5, the three degrees of freedom for T, both in the main effect and in its interactions, have been partitioned in components attributable to the linear, quadratic, and cubic terms of the polynomial. If the bulk of the sum of squares for the main effect lies in the linear term, with the quadratic and cubic terms not

significantly larger than the remainder, then it can be said that straight lines represent the data sufficiently well. If a linear component of an interaction, say $C \times T_L$, is significant, this means that the slopes of the lines of response against T are different for the two levels of C. This use of orthogonal polynomials has been described (1, 2, 9).

The interpretation of an analysis of variance such as Table 5, presents difficulties arising out of the multiplicity of tests being made at the selected significance level. Considering the half-replicate, using the ordinary variance ratio test, the interactions between C and A, C and the linear component of T, and C and S would be judged significant. If one takes account of the number of tests being made, however, Nair (12), the last mentioned, $C \times S$, would not be judged significant. The main effects of those factors involved in interactions are not of interest: those not involved in interactions are B and S and each of these is significant. If the whole replicate is considered, the interaction between C and S, which was somewhat dubious on the half-replicate, is significant. The same is true for the interaction between B and the linear component of T.

Having decided from the analysis of variance what effects to look at, the author then constructs the relevant tables of means (Table 6). All confidence limits quoted are for 95% confidence. Limits for the full replicate are naturally somewhat smaller as this uses twice the number of observations as the full replicate.

It will be noted that the half-replicate gives similar conclusions to the full replicate. It could well have been that the half-replicate was sufficiently accurate for some purposes, in which case the amount of work required would have been cut in half.

Fractional Replication with Factors at Three Levels

Finney (6)[2] has shown how the ideas just expressed can be extended to experiments with all factors at three levels. Here the three levels of the factor A are represented by (1), a, and a^2, and similarly for the other factors. Each factor will have two degrees of freedom which can be represented by symbols of the type A, A^2. Two elements are defined $a^\alpha b^\beta c^\gamma \cdots$ and $A^{\alpha'} B^{\beta'} C^{\gamma'} \cdots$ to be orthogonal if

$$\alpha\alpha' + \beta\beta' + \gamma\gamma' + \cdots = 0 \bmod 3$$

where 0 mod 3 means that on division of the sum of the product of the coefficients by 3 a remainder of zero is obtained.

To construct a one-third replicate of the 3^5 experiment, i.e., an experiment with five factors all at three levels, the alias subgroup can be selected.

$$I = ABCDE = A^2B^2C^2D^2E^2.$$

The treatment subgroup will consist of all elements orthogonal to this, namely, (1), ab^2, a^2b, ac^2, $\cdots abc$, $a^2b^2c^2$, $\cdots abc^2d^2$, $\cdots ab^2c^2d^2e^2 \cdots$.

[2]The table on page 300 of his paper is in error and has been corrected in Ann. Eugenics, 15, 276 (1950).

Table 6

Effect		Half-Replicate					Full Replicate			
B_1-B_0		3.06 ± 1.8					3.41 ± 1.11			
S_1-S_0		3.06 ± 1.8					3.16 ± 1.11			
A x C		C_0	C_1	C_1-C_0			C_0	C_1	C_1-C_0	
	A_0	36.50	42.50	6.0		A_0	36.87	42.44	5.56	
	A_1	51.87	48.75	-3.12		A_1	51.56	48.44	-3.12	
	A_1-A_0	15.37	6.25			A_1-A_0	14.69	6.00		
Confidence limits:										
for difference between two means			2.54					1.58		
for difference between two such differences			3.60					2.23		
C x T		T_0	T_1	T_2	T_3		T_0	T_1	T_2	T_3
	C_0	51.25	46.75	42.75	36.00		52.37	46.62	42.25	35.62
	C_1	48.25	46.75	45.50	41.75		48.37	46.87	45.37	41.12
	C_1-C_0	-2.75	0.00	2.75	5.75		-4.00	0.25	3.12	5.50
Confidence limits:										
for difference between two means			3.60					2.23		
for difference between two differences			5.09					3.15		
S x C		C_0	C_1	C_1-C_0			C_0	C_1	$C_1 - C_0$	
	S_0	46.75	46.12	-0.63			46.94	45.87	-1.06	
	S_1	41.62	45.12	3.50			41.50	45.00	3.50	
	S_1-S_0	-5.12	-1.00				-5.54	-0.37		
B x T		T_0	T_1	T_2	T_3		T_0	T_1	T_2	T_3
	B_0	47.25	45.25	43.00	38.00		47.12	44.87	42.37	38.12
	B_1	52.50	48.25	45.25	39.75		53.62	48.62	45.25	38.62
	B_1-B_0	5.25	3.00	2.25	1.75		6.50	3.75	2.87	0.5

There will be eighty-one elements which satisfy the orthogonality condition. The effect symbols will be confused in triplets of the type

$$
\begin{aligned}
A &= A^2BCDE = B^2C^2D^2E^2 \\
A^2 &= BCDE = AB^2C^2D^2E^2 \\
AB &= A^2B^2CDE = C^2D^2E^2 \\
A^2B &= B^2CDE = AC^2D^2E^2 \\
AB^2 &= A^2CDE = BC^2D^2E^2 \\
A^2B^2 &= CDE = ABC^2D^2E^2
\end{aligned}
$$

Two of the four degrees of freedom of the interaction between A and B are confused with two of the eight degrees of freedom of C, D, and E, but ordinarily this will not give trouble.

Data from such an experiment is given in Table 7. It seems that only one third of the usual number of runs has been made. The experiment was on a laboratory scale on a fermentation system and the five factors represent concentration of five of the ingredients. The dependent variable in the table is a function of yield. The analysis of variance is in Table 8. The only effects of importance are the interaction of factors C and E, for which the averages are given in Table 9, and the main effect of D for which the averages are given in Table 10.

Table 7

			A_0			A_1			A_2		
			B_0	B_1	B_2	B_0	B_1	B_2	B_0	B_1	B_2
C_0	D_0	E_0	63	...	...	...	...	60	...	73	...
		E_1	...	...	117	...	98	...	96	...	...
		E_2	...	130	...	110	...	...	...	...	122
	D_1	E_0	...	...	100	...	85	...	89	...	...
		E_1	...	119	...	115	...	...	...	...	141
		E_2	138	...	...	...	...	140	...	142	...
	D_2	E_0	...	92	...	106	...	...	...	...	94
		E_1	117	...	...	...	...	119	...	129	...
		E_2	...	...	154	...	154	...	158	...	...
C_1	D_0	E_0	...	...	32	...	35	...	32	...	...
		E_1	...	64	...	81	...	...	...	...	38
		E_2	100	...	...	...	...	112	...	98	...
	D_1	E_0	...	54	...	80	...	...	...	...	61
		E_1	95	...	...	...	...	95	...	95	...
		E_2	...	...	124	...	123	...	130	...	...
	D_2	E_0	60	...	...	...	...	61	...	75	...
		E_1	...	...	106	...	107	...	111	...	...
		E_2	...	130	...	147	...	...	...	...	137
C_2	D_0	E_0	...	15	...	22	...	...	...	...	0
		E_1	47	...	...	...	...	30	...	32	...
		E_2	...	...	72	...	185	...	85	...	...
	D_1	E_0	40	...	...	...	...	10	...	25	...
		E_1	...	...	69	...	69	...	60	...	...
		E_2	...	120	...	112	...	...	...	...	106
	D_2	E_0	...	...	50	...	70	...	43	...	...
		E_1	...	84	...	104	...	...	...	...	91
		E_2	123	...	...	...	...	132	...	130	...

Table 8

Source of Variance	Sums of Squares	Source of Variance	Sums of Squares
A : L	8.963	BC : LL	476.694
Q	616.395	QL	131.120
B : L	153.352	LQ	621.120
Q	323.710	QQ	375.929
C : L	23856.019	BD : LL	21.778
Q	301.488	QL	25.037
D : L	16189.352	LQ	416.148
Q	358.525	QQ	382.420
E : L	65940.167	BE : LL	110.250
Q	42.525	QL	8.898
AB : LL	84.028	LQ	290.083
QL	640.454	QQ	363.114
LQ	0.231	CD : LL	200.694
QQ	541.855	QL	528.898
AC : LL	106.778	LQ	76.676
QL	1309.037	QQ	322.003
LQ	31.148	CE : LL	2567.111
QQ	33.383	QL	625.926
AD : LL	373.778	LQ	27.000
QL	166.259	QQ	11.864
LQ	0.593	DE : LL	128.444
QQ	670.234	QL	237.037
AE : LL	26.694	LQ	120.333
QL	270.750	QQ	116.160
LQ	26.009		
QQ	179.262		
Error Sum of Squares	6147.265		
Degrees of Freedom	30		
Mean Square	204.909		
Total Sum of Squares	126582.988		

Table 9

	E_0	E_1	E_2	E_2-E_0
C_0	84.7	116.8	138.7	54.0
C_1	54.4	88.0	122.3	67.9
C_2	30.6	65.1	118.3	87.7
C_2-C_0	-54.1	-51.7	-20.4	

Table 10

d_0	d_1	d_2
72.0	94.0	106.8

It is apparent that this experiment was satisfactory. A factorial experiment was shown to be necessary as otherwise the interaction between C and E would have been completely overlooked. On the other hand, the one-third replicate was sufficiently accurate to delineate the effects with adequate precision, so the objective was obtained with only eighty-one fermentations instead of 243.

Other Possibilities in Fractional Replication

Space here does not permit an exploration of all the possibilities of fractional replication. The following, however, might be mentioned:

a. It is possible to break up all the experiments into smaller blocks so that heterogeneity in the experimental background will be eliminated from the estimation of effects. This device is known as confounding (2, 5, 6, 9).

b. In the 2^n system, for moderately large n, satisfactory designs exist which replicate smaller fractions than one half, in particular one quarter, or one eighth, etc. (2, 6, 9).

For example, with eight factors, requiring 256 runs for the full replicate, a certain set of sixty-four can be chosen to give a relatively unambiguous experiment. This is known as a quarter-replicate.

c. In the 2^n system, designs exist with some factors at four levels and others at two levels in addition to that illustrated in this paper (4).

Literature Cited

1. Anderson, R. L., and T. A. Bancroft, "Statistical Theory in Research," McGraw-Hill Book Co., Inc., New York, Chap. 16 (1952).

2. Brownlee, K. A., "Industrial Experimentation," Chemical Publishing Co., Brooklyn, N. Y., 3rd American ed. (1949).

3. Brownlee, K. A., Ann. N. Y. Acad. Sci., 52 (Art. 6), 820 (1950).

4. Brownlee, K. A., B. K. Kelly, and P. K. Loraine, Biometrika, XXXV, 277 (1948).

5. Cochran, W. G., and G. Cox, "Experimental Designs," John Wiley & Sons, Inc., New York (1950).

6. Finney, D. J., Ann. Eugen., 12, 291 (1945).

7. Fisher, R. A., "The Design of Experiments," Oliver and Boyd, Edinburgh; 4th ed. (1947).

8. Kendall, M. G., "The Advanced Theory of Statistics," Chaps. 23 and 24, II, Griffin, London (1946).

9. Kempthorne, Oscar, "The Design and Analysis of Experiments," John Wiley & Sons, Inc., New York (1952).

10. Mackay, A. C., F. Jones, and J. Dunn, Food Research, 17, 216 (1952).

11. Mood, Alexander, "Introduction to the Theory of Statistics," McGraw-Hill Book Co., New York (1950).

12. Nair, K. R., Biometrika, XXXV, 16 (1948).

A SURVEY OF THE USE OF STATISTICAL METHODS
BY A.S.Q.C. MEMBERS IN THE METALS INDUSTRY

By - Metals Technical Committee

"Enlarge the place of thy tent, lengthen
thy cords, and strengthen thy stakes."

Consideration of a biased sample of companies in the Metals Industry indicates that some statistical methods are more used than others. Such a result of course was foreseen intuitively.

The source and comment on the probable effect of the bias, description of the companies and of the individuals in those companies selected to participate in this study, the means of selecting the statistical methods included in the survey, and the measure of the extent of use or reason for not using each method comprise the present report.

Reason for making this survey.

One of the projects planned by the committee is a series of case histories of statistical methods. The selection of methods as subjects of case histories will be based on consideration of supply and demand in the Metals Industry as indicated by the results of the present survey. The companies using a given method represent the supply of possible authors or contributors to write a case history for that method. The methods that are not widely used, especially those that are not used because they are unknown, will be considered to indicate a demand for subjects of case histories.

In the belief that the information secured may be of some interest in itself, particularly to those who participated in the survey, this report is being prepared. There is currently consideration of extension of the survey to include more companies either in the A.S.Q.C. or in the Metals Industry. The committee welcomes comments or suggestions concerning such extension.

The current membership of the Metals Technical Committee, with the A.S.Q.C. Chapter affiliation and company is:

	A.S.Q.C. Chapter	Company
Frank G. Norris, Chairman	Pittsburgh	Wheeling Steel Corporation
D.H.W. Allan, Vice Chairman	Toronto	Steel Company of Canada, Ltd.
W.T. Rogers, Secretary	Cleveland	National Tube Company
Guy W. Abson	Pittsburgh	J. & L. Steel Company
Robert F. Attridge	Toronto	Steel Company of Canada, Ltd.
Marjorie M. Baskerville	Toronto	Dominion Foundries & Steel
J. R. Behrman	Philadelphia	Alan Wood Steel Company
Paul R. Brucker	Pittsburgh	Crucible Steel Company
Richard H. Ede	Chicago	U. S. Steel Company
Wm. P. Goepfert	Pittsburgh	Aluminum Company of America
Edgar H. Howells	Allentown-Bethlehem	Bethlehem Steel Company
Kenneth H. Kramer	Cleveland (Youngstown Sub-Section)	Youngstown Sheet & Tube Company

John F. Occasione	Cleveland	American Steel & Wire Division
Harold L. Springer	Pittsburgh	St. Joseph Lead Company
John V. Sturtevant	Pittsburgh	U.S. Steel Company
John W. W. Sullivan	Metropolitan	American Iron & Steel Institute

Selection of those to receive the questionnaire.

A questionnaire type of survey was used.

In general the questionnaire was sent to the highest ranking person in a given company known (or surmised) to have immediate interest in Quality Control. This individual was usually (but not always) a member of A.S.Q.C. Of the 76 replies 51 were by the one to whom the letter was addressed, and 25 were by someone to whom the questionnaire had been referred.

The obvious conclusion is that two-thirds of those responding believed that they themselves were the proper ones to reply. For the other one-third of the replies the job had been assigned.

In most instances the questionnaire was sent to someone actively engaged in, or at least intimately familiar with the use of statistical methods. This conclusion is of course a logical consequence of the method of selecting those to receive the questionnaire.

The replies classified by title of the person answering are shown in Table 1. This classification is a rough indication of the place of Quality Control in the company organization: 25 of the replies were by Metallurgists.

The next most popular title (of the respondents) is Supervisor of Quality Control.

The term Quality Control Engineer is a poor third. This job title is apparently not as popular in our sample of the Metals Industry as might be concluded from its frequent appearance in current technical literature.

The selection of the list in the above manner has had (at least) two important effects on the replies:

1. The percentage of replies was phenomenally high.
2. A bias is introduced so that the replies are not representative of the general use in the Metals Industry.

1. Extent of the replies.

The intent of the committee has been to emphasize - that the flow of information is along a two way channel and that a complete report of the assembled results would be sent to those furnishing the detailed information used to make the summary. This follow through is about the only incentive that can be offered to encourage participation. It is thought to be important.

A total of 76 replies were received from 103 questionnaires. As an auxiliary study the replies were divided into quartiles on the chronological sequence of return. No evidence of difference in the nature of the

use of statistical methods was found associated with early compared with later replies. The first two replies were from companies in Ohio. This geographical bias disappeared from the later returns.

Some letter surveys with a lower percentage of returns fail to stress one or more of the following features the combined effect of which we believe contributed to the high level of response to our survey:

1. A definite purpose of the survey (beyond mere nibbiness) that is clearly stated to the participants.
2. Distribution to a select group rather than broadside. The basis of selection of our group was such that the members were expected to have sympathy with the purpose of the survey and at least one point of contact has been established before receiving the first letter.
3. Assurance of availability of the complete report.

For the information of those interested in the mechanics of conducting this survey a copy of the letter of transmittal is appended to this report.

<u>2. Is the survey representative?</u>

The answer to this question is definitely No. The results of this sample indicate greater use of more methods than is current among the entire Metal Industry. No way is known to estimate the extent of the difference. The reason for suspecting this bias toward greater use lies in the method of selecting the companies to receive the questionnaire. It may be modestly assumed that A.S.Q.C. members are a select group. Interest and use are compatible and indeed mutually stimulating. Those most interested are most likely to belong to A.S.Q.C. Through the contacts and activities of the A.S.Q.C. come interest, knowledge and use of various methods.

<u>Description of those replying to the questionnaire.</u>

An auxiliary study furnishes some indication of the place of statistical methods in the company organization in the Metals Industry.

Much has been written about how Quality Control should be organized. As a background, it may be helpful to consider what situation actually exists, i.e., how Quality Control actually is organized.

There was some thought given to including questions designed to show the duties and training of a Quality Control Engineer. Such questions were omitted. The questions retained are believed to be factual and easy for anyone to answer objectively.

The 76 replies have been grouped by four major categories:

1. The position or job title of the individual who answered.
2. The department in which the respondent is employed.
3. The general responsibility of the respondents immediate superior.
4. The general classification of the major product of each participating company.

The answer to the question, "who replied to the questionnaire?" can

be summarized (based on the modal response). The most popular response (to each of the four above categories) was:

1. From a company engaged in iron and steel production.
2. By a Metallurgist.
3. From someone in the Metallurgical Department.
4. From someone whose immediate superior has technical responsibilities.

A temptation at this point, which some authors cannot resist, is to describe the characteristics of a mythical average reply. The fallacy of describing an "average" individual can perhaps be shown by attempting to answer a related question: "What is the expectation of finding an individual with all four modal characteristics?" There is an unknown amount of correlation among at least some of the classifications. For example we would rather expect, by intuition, to find a high number of Metallurgists in the Metallurgical Department or to find a Chief Inspector in the Inspection Department rather than in the Control Laboratory or in Research.

Neglecting the effect of this correlation we can estimate the expected number of replies by individuals who are in each of the four of the modal groups.

The probability that any one reply is from a Metallurgist is .329.

The probability that a reply is from some one in the Metallurgical Department is .342.

The probability that the one answering has a superior with a technical function is .468.

The probability that any one company is in ferrous production is .493.

The joint probability of these four conditions occurring together is:

$$.329 \times .342 \times .468 \times .493 = .026$$

For a group of 76 the expected number is:

$$76 \times .026 = 1.98 \text{ or } 2 \text{ approximately}$$

An actual check of the group discloses that of the 25 Metallurgists, 14 were in the Metallurgical Department. Of these 14 men, 10 men with a company engaged in ferrous production. Six of these 10 men reported to a superior having a technical function.

There is almost a tie between Metallurgical Department and Quality Control Department with Metallurgical slightly in the lead. It would be of interest to compare this information with that for other industries. Does some other technical function take first place corresponding to our Metallurgical Department? Or, does the Quality Control function reside in our Quality Control Department to a greater extent in other industries than it does in the Metals Industry?

The replies were grouped in three broad categories according to

respondent's superior. Slightly less than 15% of the respondent's report to someone in the operating division. The remaining 85% of the respondents was divided about equally between a technical and a strictly management function of their superiors with a technical slightly in the lead.

The nature of the methods in the questionnaire and of the responses to the use of these methods is indicated to some extent by four general statements; two pertaining to the selection of the methods and two to the companies replying.

1. No method was used by everyone.
2. Each method (but one) was used by at least one company.
3. No company used all of the methods.
4. Every company used one or more of the methods.

Selection of statistical methods.

There is no way of selecting a standard to use to judge the best set of methods to include in a study of this kind. The methods on this list were selected purely by intuition and screened and revised several times by trial. It is readily acknowledged that this list is not free from objection. It is believed, however, that moderate revision of the list would not greatly influence the general conclusions reached by this survey.

The intent was to select a broad (but not exhaustive list of methods such that:

1. Everyone replying would be using some of the methods and no one would be using all of the methods.
2. None of the methods would be used by everyone and each of the methods would be used by someone.

Three preliminary revisions of the questions and list of methods were made and circulated to members of the committee. After each response some methods and questions were dropped and some were added.

At an early stage there were several questions pertaining to organization and personnel. These were omitted on the final draft because they seemed to detract from the chief objective of the survey.

The duplication and correlation among some of the methods is recognized. This feature of the list was considered by the committee and intentional.

A more serious limitation is the ambiguity. For example, the term "Dot Chart" is used by Ezekiel to describe a chart that is quite different from the one that Wilks refers to by the same name. There is no way of telling from the replies what each method name signifies to each respondent. In order completely to remove this uncertainty it would have been necessary to accompany each method by a definition or literature reference. This possibility was considered by the committee and ruled out on the grounds of being too cumbersome, and that such a supplement to the questionnaire form would introduce more confusion than it would eliminate. The answers must therefore be interpreted in the light of what the reader believes was intended by the one who made the reply.

While recognizing the room for possible improvement in the list of

methods, it is recommended that it be retained unchanged in any extentions of the study to other companies in the immediate future.

The methods are arranged according to six broad groupings:

1. Description
2. Charts
3. Quality assurance plans
4. Correlation
5. Variance
6. Not otherwise listed (called special techniques).

Whether or not these groupings adequately cover the field of the statistical methods used in Quality Control is of course subject to discussion. Some duplication is recognized both within and among the groups. Some of the procedures that are listed as separate methods are so frequently used together that they should perhaps be either combined or partially omitted. Beyond these few explanatory remarks no attempt will be made to defend the selection of methods.

Classification of response.

The interpretation of the replies rests heavily upon the unknown points of view of the men who supplied the information. A finer classification than simply "use" or "not use" has been attempted. The division between "routine" and occasional" is of course vague and variable.

There is further, and even greater, difficulty in attempting to classify the reasons for not using a method. The reasons suggested are neither exhaustive nor mutually exclusive. Both of these limitations are reflected to some extent in the general results. There is no certainty that each one of the men indicating the same reason for not using a method, had exactly the same attitude or point of view regarding that method. Sometimes such reasons are not clearly formulated. There has been some uncertainty regarding the term, "would like to know more about." In the final analysis the committee believes that everyone who has an inquiring mind has this attitude towards all scientific methods. It has been hoped that those giving this reason for not using a method had heard of the method so that it would not be proper to say that they didn't use it because they didn't know about it, but they still felt that more knowledge or familiarity was needed before they cared to try to use it for themselves. If this interpretation accurately reflects the attitude of most of the ones giving this reply, such replies should be given considerable weight in the selection and preparation of case histories.

In view of these and other uncertainties in the interpretation of the replies the combination of replies to give an index of demand or supply is approached with caution. Recognizing that the method is cumbersome, the mode is used in reporting the results.

Comparison of methods.

The results are presented in two ways:

1. In the same order as on the survey form.
2. According to modal response.

1. Mimeographed forms were distributed - Table 2 - is a duplication of these forms. The numbers indicate the number of companies (replies) in each category for each method.
2. The replies are shown graphically for each method (Fig. 1) in order of the number of each kind of response. For example there were methods with more replies in Col. 1 (used for routine reports) than in other columns. These methods whose modal use as for routine reports are grouped together in Fig. 1.

RESULTS OF THIS SURVEY.

It is anticipated that the report of this survey will have two stimulating effects.

1. Interest in the survey.
2. Interest in the use of statistical methods.

Interest in the survey.

Even among the select group participating in this survey, some reluctance to reply was evident in this typical attitude either expressed or implied.

"One answer is not of great importance particularly if not many of the methods are being used."

It is believed that study of this report will dispel this point of view. It is important to know what methods are not being used, and why. No one person can know this, least of all the group of persons (the committee) who are somewhat familiar with most of the methods. Such information can be assembled only by the intelligent co-operation of a large group. Any reader who applies to the Chairman of the Metals Technical Committee will be furnished two copies of the blank forms (one for his file). If a sufficient number of these are returned, a supplement to this report will be issued covering these replies.

Interest in the use of statistical methods.

As stated earlier, the primary purpose of this survey is to furnish information for the selection of case histories of statistical methods used in the Metals Industry. The purpose of case histories is frankly promotional. It is in line with the first stated purpose of the American Society for Quality Control: "To create, promote, and stimulate interest......"

A familiar limitation in accumulating data is that some characteristics are changed by the process of taking a measurement. This limitation is believed to apply in attempting to determine the use of statistical methods. Certainly the questionnaire itself will stimulate interest. In fact such stimulation is a secondary intent of the project. Presumably use also will increase. It is difficult to devise a method to confirm the results.

Comparison of methods by general groupings.

The 83 methods have been rather arbitrarily grouped into six

categories each on a separate page of the questionnaire form. An index of use has been computed that is intended to show the extent of use (both routine and occasional) of these groups of methods. The scale of the index extends from 0 to 1. A value of 0 would indicate that no method on a given page is used by any company. A value of 1 would indicate that all of the methods were used by all of the companies. A value of .25 would indicate that all of the methods were used by one-fourth of the companies, or that one-fourth of the methods were used by all of the companies, or that one-half of the methods were used by one-half of the companies.

The routine use ratio is also shown. This value is the ratio of routine use to total use.

The descriptive methods are those that are most used, and the special techniques the least used. These results are about as expected.

The wide extent of the use of correlation, (mostly for occasional studies) may be unexpected by some. Recall that the early work by Chancellor emphasized the value of correlation. A still earlier application of least squares methods to Metallurgical problems were studies of the effect of composition on tensile strength. The current popularity of correlation and regression methods probably reflects a logical development from such a background and is also a testimony to the general value of the results of these methods over a long period of time.

Note that more companies are using more charts with limits than without limits. For routine use, charts with limits are more than twice as popular as those without limits. This response suggests adoption of control charts as distinguished from simply graphical plotting of data or trends.

The ratio of routine to total use is the highest for quality assurance plans and is the lowest for analysis of variance.

Based on the indication of the routine use ratio it is concluded that analysis of variance, correlation and special techniques are mostly used for occasional studies rather than for routine reports or studies. These three groups are definitely the least used as a matter of routine. The order of popularity among the three is much less certain, as this distinction is based on the small differences among the ratios .120, .158 and .169 which are considered as indicating about the same general level.

Quality assurance plans are distinctly routine in use as indicated by the routine use ratio of .634 which is the highest of 7 groups.

Charts with limits and descriptive statistics are each at the same general level (.493 and .478 respectively). Charts without limits are used more for occasional than for routine problems as indicated by the routine use ratio of .295.

Another way of comparing both the individual methods and the general groups is based on the following arbitrary assignment of point values:

Routine use	5
Occasional use	3
Has been discontinued	1
Do not know about	-1

The eight methods with the highest point value are:

Point Value	Method
406	Chart of averages - small groups
330	Average
318	100% Inspection
290	Chart of range
284	Chart of averages - large groups
283	Range
273	Chart of Individual Values
253	Standard Deviation

The method with the lowest point value was serial quadrant sum with - 34.

Group	Average Point Value	Highest Point Value	Lowest Point Value
Description of Samples	156	330	27
Charts	146	406	24
Quality Assurance	97	318	19
Correlation	85	142	13
Variance	46	109	9
Special Techniques	6	134	34

The wide spread within each of these groups suggests that no one group has a monopoly on popularly used methods.

In general this summary is in agreement with other means of showing the results of the survey.

This portion of the report is intended to be suggestive of several possible means of analyzing the results. There is no claim that any one of these comparisons is absolutely the best way of ranking the statistical methods used in the Metals Industry.

Most useful methods.

At the bottom of each page was the question, "which of the above methods do you consider the most useful?" The answers comprise an opinion survey.

The most popular method (voted among the most useful by the most companies) in each group is as follows:

Group	Method	No. of Votes
Description of Samples	Average	49
Charts	Average of Small Groups	32
Quality Assurance Plans	100% Inspection	32
Analysis of Variance	Single Criterion	15
Correlation	Linear Regression (one independent variable)	22
Special Techniques	t Test	33

The 65 methods mentioned by one or more companies as being most

useful are distributed among the six general groups as follows:

Description	7
Charts	16
Quality Assurance	8
Variance	10
Correlation	10
Special Techniques	14
Total	65 Methods

The following methods were voted as most useful (within a given page group) by 15 or more companies.

The way the questions were presented (one at the bottom of each page) produces a stratified sampling of the methods. Each page is certain to be represented. The results show the respondents opinion of the most useful method within each group. There has been no effort to estimate the relative usefulness among the general groups. In fact it is rather difficult to compare usefulness of methods unless they serve the same general purpose. This difficulty exists to some extent even among methods within the same general group.

Method	Indicated Most Useful by This Number of Companies
Description	
Averages	49
Range	21
Standard Deviation	18
Table of Grouped Frequency	19
Charts	
Averages (large groups)	18
Averages (small groups)	32
Range	21
Quality Assurance	
Linear Correlation Coefficient	19
Multiple Linear Regression	15
Graphical Methods	15
Special Techniques	
t Test	33
F Test	18

Opportunity was provided to write in other methods that are used. The other methods mentioned are:

Quality Assurance Plans.

Routine use:

1. Time Sequence Sampling.

2. Screening small lots on basis of first sample.
3. A N Specifications.

Design of Experiments.

Occasional Use.

1. Snedecor method as modified by Ede.
2. Observation by Classification.
3. Analysis of Variance for Vectors.

Correlation.

Routine use.

1. Observation.
2. Grouped correlation.

Occasional Use.

1. Graphical multiple correlation.
2. Serial correlation for time, series, analysis correlogram.
3. Wherry - Doolittle technique.
4. Non-formal Data Grouping.
5. Correlation based on minimum squares of othogonal deviation.

There is evidently some duplication. Those who gave this response must have felt that they had a use that was sufficiently different to justify separate mention.

It is believed that the 80 methods listed and the 13 methods that were indicated in the space provided for, "other methods" are a good indication of the methods currently being used in the Metals Industry.

Only one of the methods - serial quadrant sum was used by no one of the companies responding to the questionnaire.

Three other methods, although used by a few companies, were unknown to more companies than was the serial quadrant sum. These were:

Flow Channel
Linear estimate of sigma.
Discriminant functions.

Application of the Survey to the Selection of Case Histories.

For the purpose of selecting methods to use in the preparation of case histories the replies were considered in a slightly different way. Effort has been made to combine the results so as to obtain a use index for each method. It is realized that this treatment is based on assumptions that are purely arbitrary. It does have the appeal of simplicity and has been thought to be helpful focusing attention on various methods.

A demand factor is shown by the answer; "A method is not used because it is not known" or, "because we would like to know more about

it." The number of companies that are using a given method for either type of report - routine or occasional, indicates the extent of possible sources of case histories.

The sum of the number of companies using a given method is termed the supply factor.

The following paragraph and Table 3 are extracts from the sub-committee report prepared by W. P. Goepfert.

"The decision for inclusion of the subject in the attached tabulation was based on a supply index of not less than 5 and a demand index of not less than 15. These indexes were determined as indicated in the column headings of the tabulation. This approach was used as a simplification and the criteria of inclusion merely represents what appeared to be a reasonable way of separating out the more desirable subjects. Column 7 is included in the Demand group merely as additional supporting evidence of the demand criteria although it is not a part of the index."

Summary.

1. The general pattern of the responses is in agreement with the results expected. This agreement is interpreted as a rough indication of the validity of the sample and the accuracy of the replies.
2. All of the obvious sources of bias are in the direction expected to inflate the reported use. Therefore the use of these methods by a larger or more general population is believed to be not more than that shown by this limited sample.
3. Tentative plans have been made to extend this survey - using essentially the same questionnaire - to a larger sample of the Metals Industry.
4. The results are thought to be an interest in themselves. They will be used as a guide in the selection and preparation of case histories of methods used in the Metals Industries.

Appendix.

Mechanics of Assemblying the Data.

The procedure used to assemble the information from the returned questionnaires is briefly described. Mechanical equipment of whatever nature cannot do anything that cannot be done by manual methods. The advantage machine sorting is that much less time is required and errors due to carelessness or fatigue are eliminated. There is no claim that the methods we used are the best possible methods for this application. We simply used the equipment available to do the job to be done.

Two card forms were used.

One card of the first form was punched for each company showing the reply of that company for each method. This part of the job would have been difficult if not impossible with fewer columns available. Eighty six card columns were used. Three were for company code. There is one column for the use code for each of the 83 methods.

When it was decided to assemble the information 76 replies had been received. The 76 cards were sorted on each method column. The numbers

of cards falling into eacn sorter pocket was recorded manually.

Using this summarized information card form 2 was punched. There is one card for each method. This is essentially the information distributed as the preliminary report.

Several runs were made on the tabulator with the entire deck of methods sorted by use.

This deck was also used in arranging the methods by modal use.

Table 1 - Classification of Replies to Questionnaire

1. Response by Position	Percentage
Metallurgist	32.9
Statistician	9.2
Quality Control Engineer	11.8
Chief Inspector	9.2
Supervisor of Quality Control	22.1
Operations Researcher	2.6
Others	11.8
2. Response by Departments	
Quality Control	27.4
Metallurgical	34.2
Control Laboratory	2.7
Inspection	13.7
Research	11.0
Methods	1.4
Statistical	4.1
Others	5.5
3. Response by Product	
Ferrous Production	49.3
Ferrous Casting	13.7
Ferrous Fabrication	11.0
Non-Ferrous Production	12.3
Non-Ferrous Fabrication	11.0
Others	2.7
4. Response by Superior	
Management	39.0
Operating	14.3
Technical	46.8

AMERICAN SOCIETY FOR QUALITY CONTROL
September 30, 1953.

SURVEY OF STATISTICAL METHODS
IN THE METALS INDUSTRIES
BY METALS TECHNICAL COMMITTEE
OF A.S.Q.C.

The survey of statistical methods by the Metals Technical Committee of the American Society for Quality Control has yielded responses from 76 of the 103 Companies to which questionaires were sent.

A summary of these 76 replies is attached and is being sent to all who received the questionaire. Additional copies are available on request.

This summary is a duplicate of the questionaire form showing the number of Companies either using or not using each methods. The nature (routine or occasional) of the use, or the reason for not using the method is indicated by the various column headings. The column headed 0 (first column at right of page) indicates the number of Companies making no comment regarding a given method. For example: 59 Companies use averages for routine reports, 12 use averages for occasional reports, 2 Companies do not use averages, one because the method is unknown and one because it does not fit their needs. Three Companies omitted comment regarding their use of averages.

Further analysis of the replies is planned and will be the subject of a more detailed report.

The Committee thanks you for your participation in this survey and will welcome suggestions or comments regarding this project.

Frank G. Norris.

Frank G. Norris, Chairman
Metals Technical Committee.

FGN:lmm

TABLE 2

1 DESCRIPTION OF SAMPLES

We now use:
a. For routine reports,
b. For occasional reports or problems.

We do not now use: (Check reason)
Do not know about — c
Would like to know more about — d
Does not fit our needs — e
Has been discontinued — f
Plans to use if needed — g
Other (Specify) — h

a	b		0	h	g	f	e	d	c
59	12	1. Average	3	0	0	0	1	0	1
8	27	2. Median	8	0	15	2	16	0	0
2	23	3. Mode	12	0	19	0	15	2	3
6	12	4. Percentile (Including Quartiles)	15	0	14	0	20	2	7
44	21	5. Range	5	0	4	0	0	2	0
32	31	6. Standard Deviation	7	0	4	0	0	2	0
0	17	7. Tests for skewness (Specify what test)	11	1	13	1	23	6	4
0	10	8. Test for kurtosis (flatness)	13	1	17	1	23	7	4
26	31	9. Table of grouped frequency distribution.	5	0	4	0	5	2	3
14	25	10. Table of cumulative frequency distribution	11	0	14	0	7	1	4

We now use:
With control limits:
a. For routine charts.
b. For occasional use.
Without limits:
A. For routine charts.
B. For occasional use.

We do not now use: (Check reason)
Do not know about ____ c
Would like to know more about ____ d
Does not fit our needs ____ e
Has been discontinued ____ f
Plan to use if needed ____ g
Other (specify) ____ h

a	b	A	B		0	h	g	f	e	d	c
17	14	19	21	1. Averages of large groups.	10	0	5	0	3	1	1
41	21	18	16	2. Averages of small groups.	5	0	1	0	0	0	0
8	18	8	31	3. Grouped frequency distribution.	7	0	6	0	3	1	1
2	12	4	24	4. Cumulative frequency distribution on linear scale.	13	0	10	1	9	4	4
3	2	0	13	5. Cumulative frequency distribution on a distorted scale (arithmetic probability)	13	1	9	2	14	9	9
2	2	0	8	6. Cumulative frequency distribution on a distorted scale. (logarithmic probability)	14	1	8	3	19	11	10
8	13	3	24	7. Averages showing spread of the group about the average.	9	0	8	0	8	3	3
3	4	0	9	8. Mid- range (max.+ min.)/2	12	2	14	1	22	4	4
33	20	6	12	9. Range	8	1	7	0	0	1	1
6	16	5	16	10. Standard deviation (of large groups).	10	0	14	1	10	1	1
9	20	3	13	11. Standard deviation (of small groups).	9	0	12	2	9	1	1
23	14	13	17	12. Individual Values.	12	0	6	1	4	0	1

10	12	11	10	13. Moving Averages	10	0	12	1	5	4	4
6	8	2	8	14. Moving Ranges	11	1	18	3	10	9	4
4	3	0	4	15. Runs above and below median.	12	1	17	0	13	14	8
5	2	1	2	16. Runs up and down.	12	1	16	0	15	14	18
4	4	0	6	17. Medians.	13	1	20	0	22	4	3
8	16	2	6	18. c Chart.	12	0	16	1	10	5	2
6	10	3	3	19. c/u Chart (defects per unit).	11	1	14	2	16	7	3
17	14	8	8	20. p Chart (including pn)	9	1	10	0	10	5	1
7	3	1	4	21. Demerits (Weighted defects)	14	2	16	0	17	10	5

3 QUALITY ASSURANCE PLANS

We now use:
a. For routine reports.
b. For occasional reports or problems.

We do not now use: (Check reason)
Do not know about
Would like to know more about
Does not fit our needs
Has been discontinued
Plan to use if needed
Other (specify)

a	b				h	g	f	e	d	c
56	13	1.	100% Inspection	0 3	0	2	1	1	0	0
25	16	2.	Single Sampling	11	1	7	3	10	2	1
			DOUBLE SAMPLING							
6	11	3.	Dodge-Romig	15	1	13	7	16	3	4
14	8	4.	Mil. Std. 105A	14	3	13	2	16	2	4
2	4	5.	Navy Tables	18	5	14	3	21	3	6
5	1	6.	Other (specify)	48	2	2	3	10	2	3
11	10	7.	Sequential Sampling (Including Multiple Sampling)	14	1	15	1	18	4	2
5	7	8.	AOQL Plan for Continuous Production	13	1	16	1	22	8	3
18	8	9.	Sampling Inspection by Variables	12	2	13	1	12	7	4
4	6	10.	Lot-Plot	14	3	11	1	24	9	4

4 ANALYSIS OF VARIANCE

We now use:
a. For routine reports.
b. For occasional reports or problems.

We do not now use: (Check reason)
c. Do not know about
d. Would like to know more about
e. Does not fit our needs
f. Has been discontinued
g. Plan to use if needed
h. Other (specify)

a	b		O	h	g	f	e	d	c
4	32	1. Single Criterion	13	0	3	1	5	10	8
3	29	2. Double Criteria	15	0	3	0	5	13	8
4	25	3. Three or more criteria	14	0	4	0	5	12	11
1	14	4. Co-variance	14	1	9	0	5	18	14
2	6	5. Unweighted means in a k by m by n classification	18	2	9	1	4	16	18
3	10	6. Unweighted means in a 2 by 2 by 2 classification.	18	1	8	1	4	11	19
		DESIGN OF EXPERIMENTS							
1	16	7. Factorial Design	15	2	11	0	6	14	11
1	6	8. Incomplete Block	17	3	15	0	6	14	11
1	11	9. Randomized Blocks	17	3	11	0	6	15	12
1	5	10. Other (Specify)	50	2	2	0	3	5	8

5 CORRELATION

We now use:
a. For routine reports.
b. For occasional reports or problems.

We do not now use: (Check reason)
c. Do not know about
d. Would like to know more about
e. Does not fit our needs
f. Has been discontinued
g. Plan to use when needed
h. Other (specify)

a	b		O	h	g	f	e	d	c
7	37	1. Linear regression (one independent variable)	10	0	3	0	4	6	9
7	38	2. Linear correlation coefficient.	9	0	3	0	4	9	7
4	18	3. Partial correlation coefficient.	14	0	9	3	6	11	11
4	27	4. Multiple Linear regression.	13	0	6	0	4	11	11
4	27	5. Multiple correlation coefficient.	13	0	7	0	3	11	11
2	19	6. Beta coefficient.	13	0	3	0	8	16	15
1	19	7. Non-linear regression	11	0	10	4	4	14	13
1	17	8. Non-linear correlation coefficient.	11	0	10	6	4	14	13
13	33	9. Graphical methods.	10	0	1	2	1	8	8
2	4	10. Other methods (Specify)	55	0	1	0	2	3	9

6 SPECIAL TECHNIQUES

We do not now use: (Check reason)
Do not know about — c
Would like to know more about — d
Does not fit our needs — e
Has been discontinued — f
Plan to use if needed — g
Other (specify) — h

We now use:
a. For routine reports.
b. For occasional reports or problems.

a	b		O	h	g	f	e	d	c
10	33	1. t Test	12	1	2	0	1	2	15
6	34	2. F Test	13	1	0	0	1	5	16
0	5	3. Sequential t Tests.	20	2	11	0	3	13	22
1	6	4. Power of t Tests.	19	2	8	0	4	14	22
1	1	5. Eta (Mean square successive difference).	19	1	12	0	3	13	26
0	2	6. p λ (λ test for gross errors).	17	1	5	0	3	17	31
0	1	7. p (g) (For an extreme variance).	17	1	6	0	3	16	32
3	8	8. Bartlett's test. (For homogeneity of variance)	18	2	10	0	2	9	24
1	5	9. Quadrant sum.	17	1	5	0	4	11	32
0	0	10. Serial quadrant sum.	18	1	6	0	3	14	34
0	1	11. Lag quadrant sum.	17	1	6	0	3	14	34
1	9	12. Spearman's Rank Correlation. (d)	17	2	4	1	4	10	28
1	4	13. Kendal's Rank Correlation (tau)	17	1	8	0	5	10	30

4	24	14. Chi Square (general)	13	1	9	1	3	6	15
3	21	15. Chi Square to test normality.	14	1	10	1	4	7	15
4	12	16. Chi Square to test a 2 by 2 classification.	17	1	9	1	3	9	21
0	11	17. Plot on Binomial probability paper.	14	2	13	0	7	11	18
		PROBIT ANALYSIS							
0	1	18. a. As designed (for sigmoid dosage-response).	19	1	3	0	9	10	33
0	3	19. b. Modified Use.	18	1	4	0	6	11	33
1	3	20. Flow Channel.	16	1	2	0	3	10	40
1	2	21. Linear Estimate of Sigma.	17	2	4	0	4	9	37
1	1	22. Discriminant functions.	19	2	5	0	5	8	35

Table 3 Topics Recommended for Case Histories Based on Supply and Demand

Subject	Supply			Demand			
	Routine Reports	Now Used For Occasional Reports	Supply Index (1 + 2)	Do not know about	Like to know more	Demand Index (4 + 5)	Does not fit need
	1	2	3	4	5	6	7
Linear Regression 1 independent variable	7	38	45	7	9	16	4
Demerits (Weighted defects)	8	7	15	5	10	15	17
Linear Correlation Coefficient	7	38	45	7	9	16	4
Graphical methods (of Correlation)	13	33	46	8	8	16	1
t Test	10	33	43	15	2	17	1
Analysis of Variance Single Criterion	4	32	36	8	10	18	5
Arithmetic probability	3	15	18	9	9	18	14
F Test	6	34	40	16	5	21	1
Analysis of Variance Double Criteria	3	29	32	8	13	21	5
Chi Square (general)	4	24	28	15	6	21	3

Logarithmic probability	2	10	12	10	11	21	19
Chi Square to test normality	3	21	24	15	7	22	4
Multiple Linear Regression	4	27	31	11	11	22	4
Multiple Correlation Coefficient	4	27	31	11	11	22	3
Partial Correlation Coefficient	4	18	22	11	11	22	6
Runs above and below median	4	7	11	8	14	22	13
Analysis of Variance 3 or more Criteria	4	25	29	11	13	23	5
Factorial Design of Experiments	1	16	17	11	14	25	6
Non-linear Correlation Coefficient	1	17	18	13	14	27	4
Non-linear Regression	1	19	20	13	14	27	4
Randomized Block	1	11	12	12	15	27	6
Incomplete Block	1	6	7	14	14	28	6
Plot on Binomial Probability Paper	0	11	11	18	11	29	7
Chi Square to Test 2 x 2 Classification	4	12	16	21	9	30	3
Unweighted Means 2 x 2 x 2 Classification	3	10	13	19	11	30	4

Beta Coefficient	2	19	21	15	16	31	8
Runs Up & Down	6	4	10	18	14	32	15
Co-Variance	1	14	15	14	18	32	5
Bartlett's test (For homogeneity of variance)	3	8	11	24	9	33	2
Sequential t Tests	0	5	5	22	13	35	3
Power of t Tests	1	6	7	22	14	36	4
Spearman's Rank Correlation	1	9	10	28	10	38	4
Quadrant sum	1	5	6	32	11	43	4

Table 4 - Comparison of Use by General Groupings

Method Group	Index of Use	Routine Use Ratio
Description of Samples	.526	.478
Correlation	.374	.158
Quality Assurance Plans	.303	.634
Charts - with limits	.283	.493
Variance	.230	.120
Charts - without limits	.227	.295
Special Techniques	.135	.169

Appendix 2

Metals Technical

Address reply to:
Frank G. Norris
Wheeling Steel Corp.
Steubenville, Ohio

April 28, 1953

Dear Sir:

The Metals Technical Committee of the American Society for Quality Control is studying modern statistical methods that are used in research, development, production and inspection in the metals industries.

The study is intended to reveal methods that are advantageously used There is a need for more published information on such methods. At a later time, those methods will be described by means of case histories and published by the Society, as a guide to the planning of research and development work and to the control of quality of incoming material and manufactured product in the metals industries.

The survey outlined in the accompanying sheets is the first step in the program of publishing the case histories. One of the important objectives of this survey is to locate the need and interest in methods, that are not now generally used. For this reason, some of the less known methods are included in the list. To indicate that you do not know about a method, or that you would like information about the possible use, will be particularly helpful in attaining this objective. Obviously such information must come from a general survey and cannot possibly be provided by a few committee members.

This letter is being sent to a limited number of companies. The results of the survey will be sent to you if you request them. On the basis of results obtained, the survey will be extended. The results of the extended survey also will be sent to you if you request them.

I should appreciate you having the attached survey sheets completed and returned to me. One set is intended for your file.

I want to thank you for your cooperation on the survey.

Yours very truly,

Frank G. Norris

Frank G. Norris,
Vice Chairman

FGN:jlm

Address reply to:
Frank G. Norris
Wheeling Steel Corporation
Steubenville, Ohio

June 15, 1953

Dear Sir:

On May 8, a questionnaire was sent to you regarding the Survey of Statistical Methods in the Metals Industries, by the Metals Technical Committee of the American Society for Quality Control.

As yet, the reply from your company has not been received, possibly due to the time required to prepare and transmit the answer or because of the urgency of other matters. It is quite possible that your reply is already in the mail, in which case, you will, of course, disregard this follow-up.

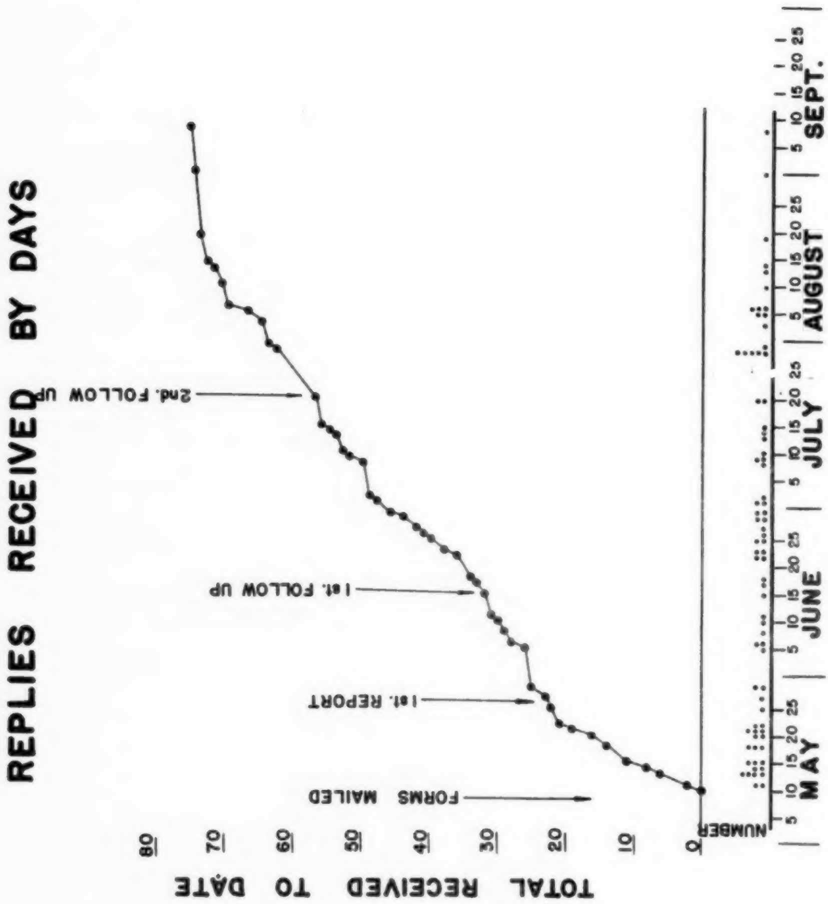
If this delay has been due to need for further information or explanation, will you please bring this need to my attention?

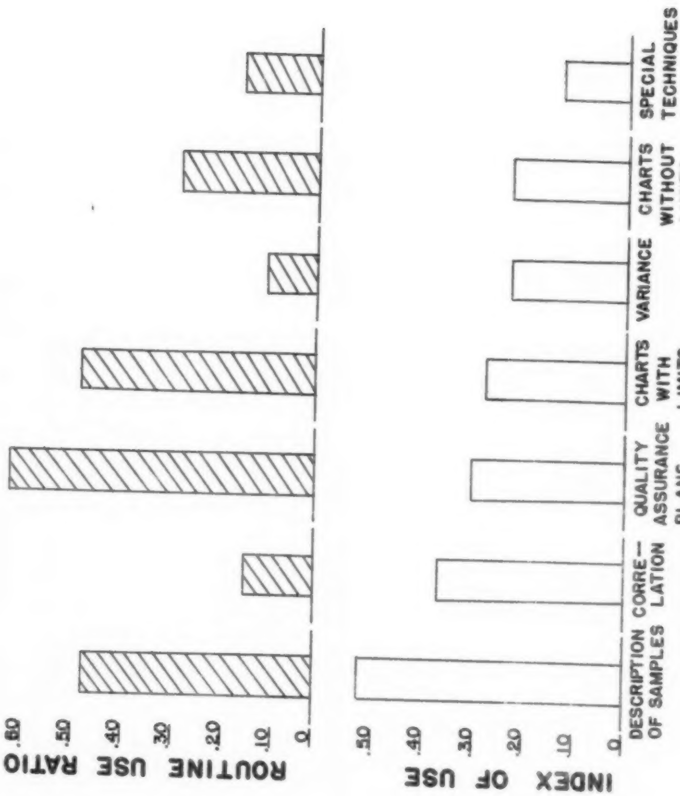
Thank you for your interest and help with this survey.

Yours very truly,

Frank G. Norris
Chairman
Metals Technical Committee

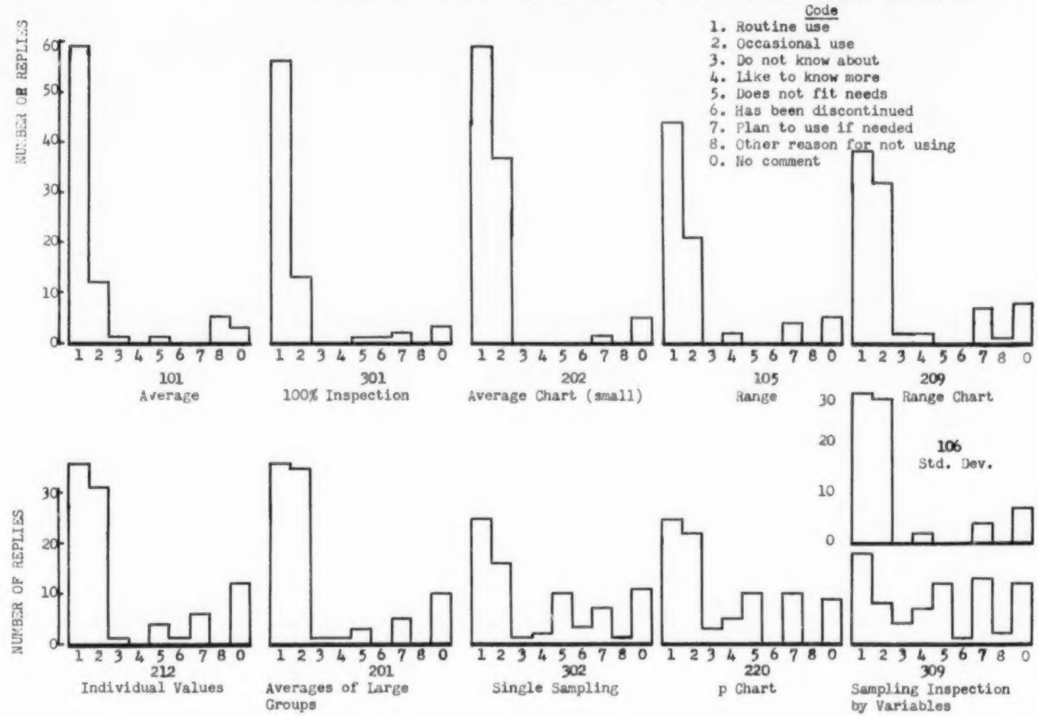
FGN:jlm


REPLIES RECEIVED BY DAYS
TOTAL RECEIVED TO DATE
0
10
20
30
40
50
60
70
80
FORMS MAILED
1st. REPORT
1st. FOLLOW UP
2nd. FOLLOW UP
NUMBER
MAY
JUNE
JULY
AUGUST
SEPT.

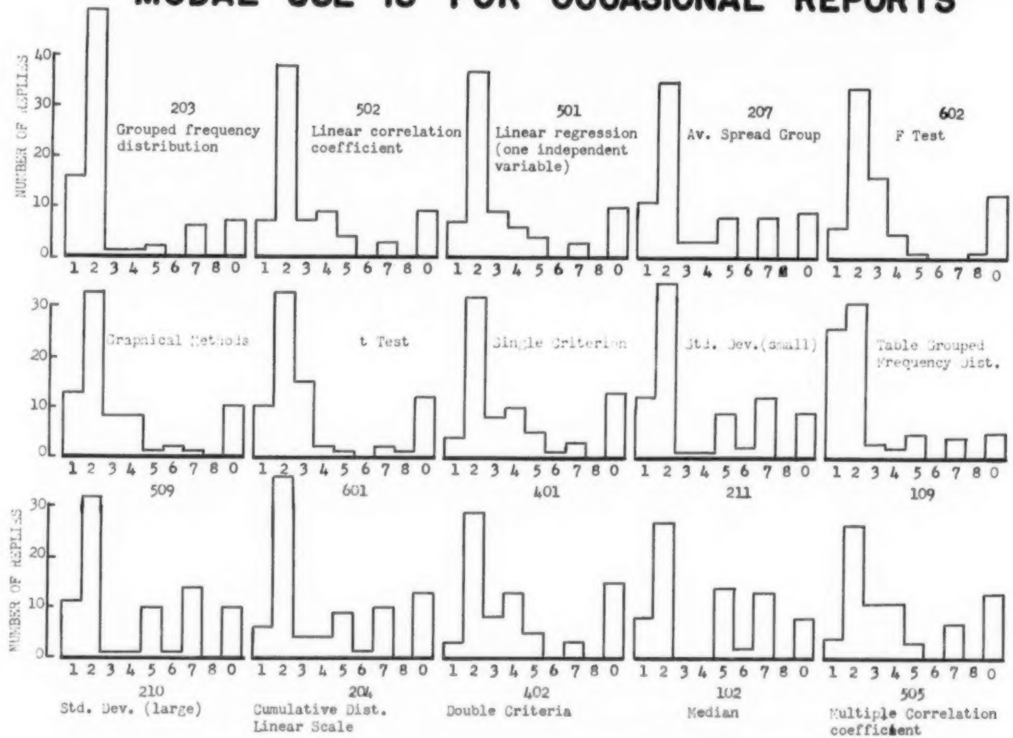


ROUTINE USE RATIO
.60
.50
.40
.30
.20
.10
0
INDEX OF USE
.50
.40
.30
.20
.10
0
DESCRIPTION OF SAMPLES
CORRE- LATION
QUALITY ASSURANCE PLANS
CHARTS WITH LIMITS
VARIANCE
CHARTS WITHOUT LIMITS
SPECIAL TECHNIQUES

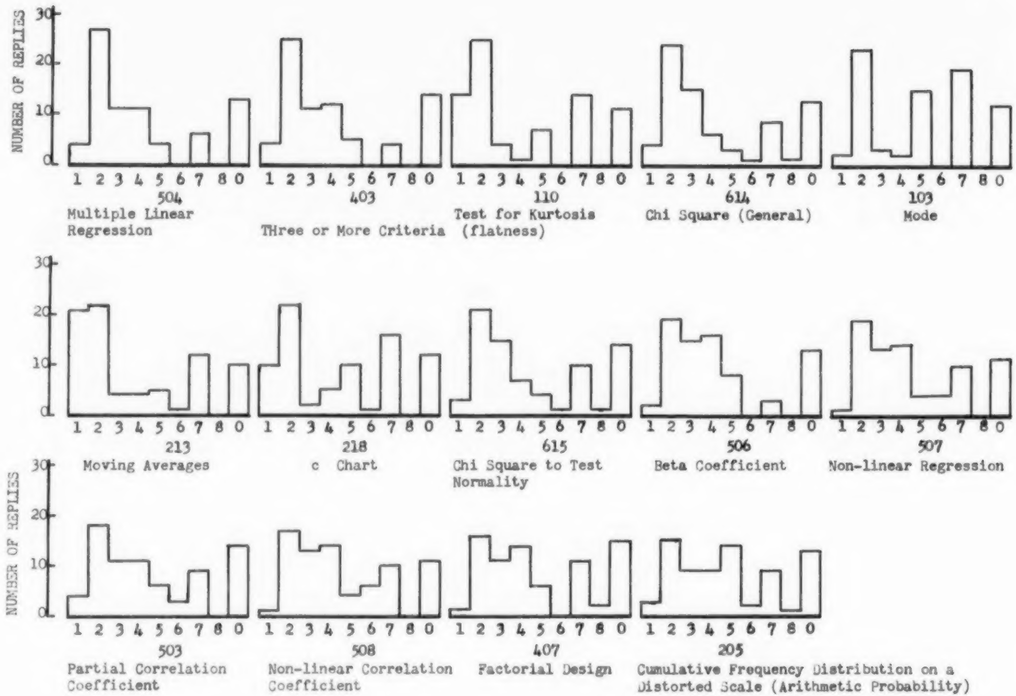
MODAL USE IS FOR ROUTINE REPORTS



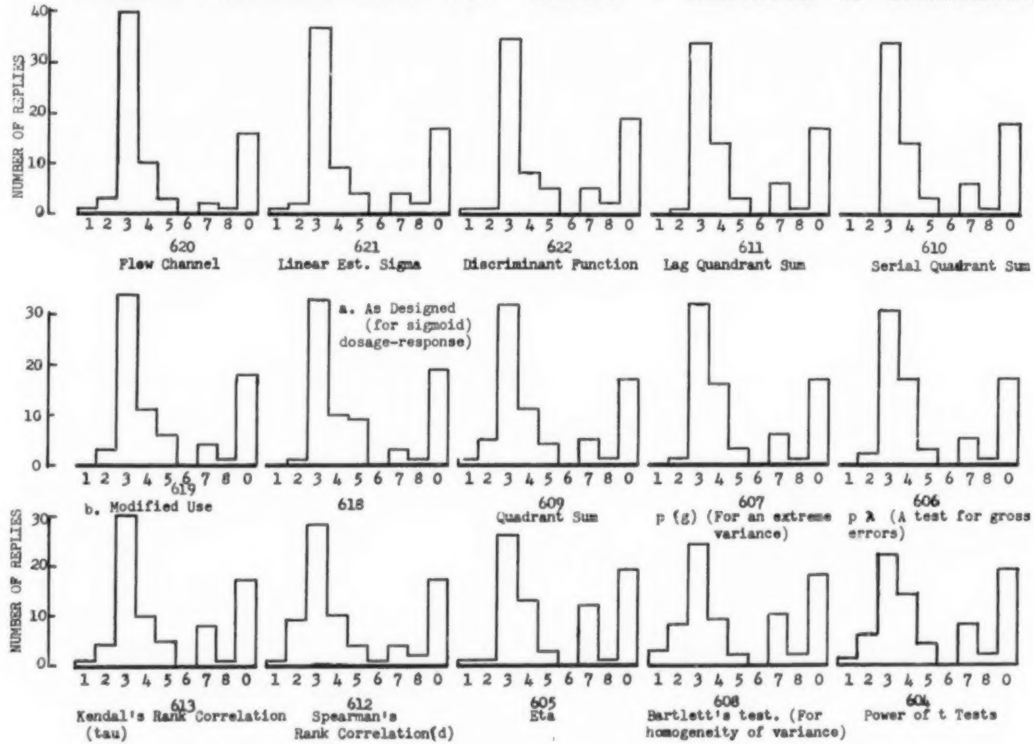
MODAL USE IS FOR OCCASIONAL REPORTS



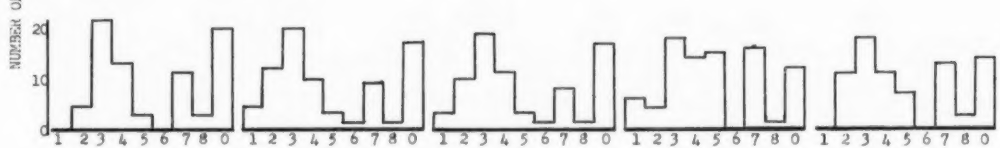
MODAL USE IS FOR OCCASIONAL REPORTS



MODAL REASON FOR NOT USING — METHOD IS UNKNOWN



MODAL REASON FOR NOT USING METHOD IS UNKNOWN



603
Sequential t Tests

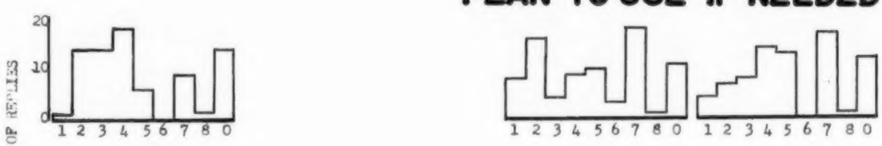
616
Chi Sq. to Test 2 by
2 Classification

406
Unweighted Means
2x2x2 Class

216
Runs Up and Down

617
Plot on Binomial
Probability Paper

PLAN TO USE IF NEEDED

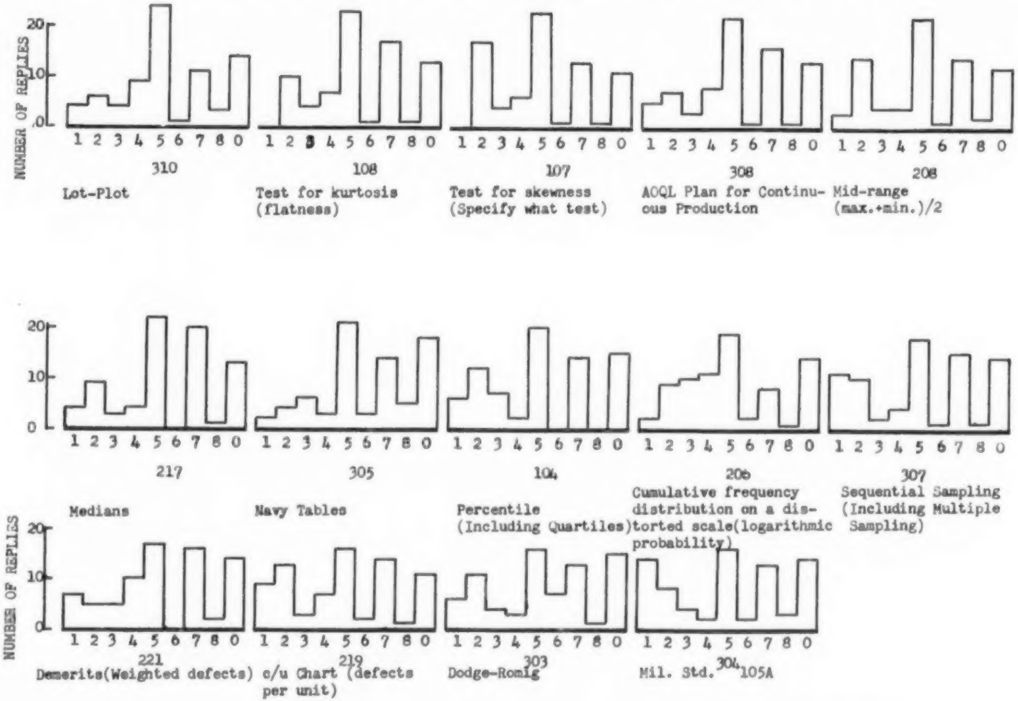


404
Co-Variance Would Like
to Know More About

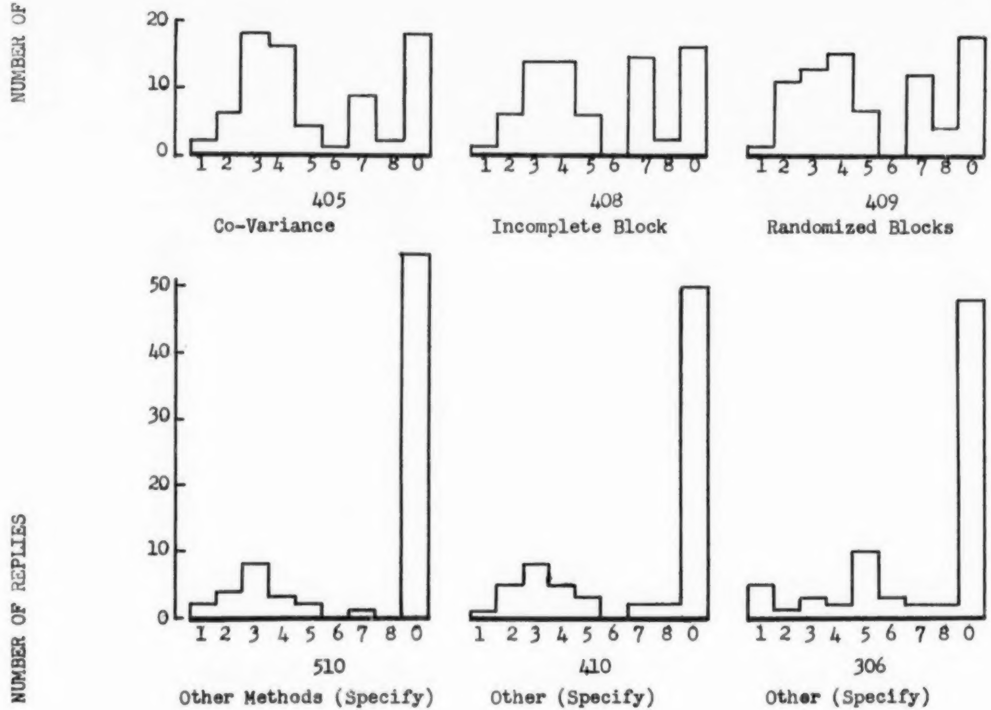
214
Moving Ranges

215
Runs Above and Below
Median

MODAL REASON FOR NOT USING— DOES NOT FIT NEEDS



MODAL REPLY NO COMMENT



CAN INCENTIVES AND QUALITY LIVE TOGETHER?

Wayne M. Biklen
American Safety Razor Corporation

This topic, "Can Incentives and Quality Live Together?", has been the subject of heated debate and long hours of discussion pro and con for the past fifteen to twenty years. Time Study Engineers, advocates of Scientific Management techniques, and experts in the field of Industrial Engineering and Management, perhaps far more experienced than this writer, have developed various solutions for this problem, and to be sure, the future will undoubtedly bring forth further valuable data to throw more light on this subject.

This writer, in no way refutes those experts, nor does he claim to be an expert, but merely enumerates here his ideas in light of almost twenty years experience in multi-plant mass production operations. He fully realizes that the thoughts expounded here will provoke further discussion.

However, it is his opinion that a "marriage" between modern forceful Quality Control and our concept of Incentives in mass production is quite possible. Not only (in his opinion) is this "marriage" desirable, but mandatory to achieve optimum results, with regard to productivity, quality and costs.

Basically, we must admit, for industry to be successful today, in face of higher material and labor costs, in face of stiffer competition, it must explore to the fullest extent and capitalize on all proven techniques available regarding the "Five M's".

1. Management
2. Men
3. Machines
4. Material
5. Methods

(Some will argue this may be resolved into three "M's", but to plug all loopholes, and prevent any false assumptions, we are listing them as above.)

We may pass over the first four "M's", with a minimum of discussion. To be sure, in the fiercely fought arenas of competitive American industry today, Management must be ever alert in every phase of operations; accounting, engineering, budgeting, personnel relations, manufacturing, movement of materials, cost-reduction systems, planning, scheduling, use of automatic devices, etc., - in short - Management must constantly be "on its' toes".

Likewise, under the category of "Men", our most valuable ingredient in operating this highly complex, but successful, mass production system, we must have people,

men (or women) wo are trained and in possession of the know-how. We must select men, teach them, give them adequate tools, keep them completely informed, - in short - they are human beings, thus, treat them so!! Without men, properly selected, trained, oriented, trusted, kept informed, and respected, this way of doing things, this technical know-how for mass producing would deteriorate. Progressive industry today would not dream of permitting its machinery to fall into dis-repair. The effect upon production becomes immediately evident when a machine fails. All too often, however, the role of the man as an integral part of a process is overlooked. Progressive Management, alerted by Quality Control, must lead the way in recognizing this important factor and help prevent the man from falling into dis-repair. This implies his psychological well-being as well as his operating ability.

Under the third "M", "Machines", we are agreed that in order to do the best job, equipment must remain efficient, modern, replaced as required, and that vigilence constantly be maintained regarding improvements, automatic devices, etc.

Under "Materials", we might enlarge somewhat as compared to the above, in that, in our opinion, very few Statistical Quality Control programs can be successful, if materials are not received, under a controlled process using statistical techniques. That is, if we expect to maintain correct, intelligent, workable incentives, - materials, components, assemblies or what have you, purchased from outside, or received from another department or division, must be received with the quality level known. It is the writer's firm belief, the future will see more and more vendees demanding of vendors, that they not only show evidence of using modern Statistical Quality Control in their process, but that in each and every shipment, certification be given regarding the shipped material's Quality Level, either through Acceptance Sampling procedures, or submission of control charts for variables. Only by these devices, will the department receiving this material, or the plant purchasing these items be able to intelligently know what to expect from subsequent departments processing these materials, quality-wise. And - how can one expect quality workmanship, quality of product, through incentives, if the incoming quality level is not known? In other words, if any incentive system is to be successful in relation to quality, the quality levels (or percent defective) of material or items being purchased for processing, or being sent from one area of the plant to another for further processing, must be known, or determined preferably by Statistical Quality Control. Too often this point is either missed by Management given low priority, or the quality of material is assumed or guessed to be within limits. Sound factual evidence, unbiased, is required and Statistical Quality Control techniques are the best means for determining these quality levels, as a prerequisite for intelligently correlating Incentives with Quality.

This discussion is intended to revolve around the fifth

"M", "Methods", as they fit into the incentive picture with Statistical Quality Control.

To digress for a moment, let us go back to the 1800's, wherein American ingenuity slowly, but steadily transformed hand selective, painful, slow, assembling into mass random assembling. We would have witnessed kitchens and barns originally used by the entrepreneur or partnership for producing a few products daily. Gradually, one and two story buildings were utilized; with the advent of electricity, line-shafts, jack-shafts, counter-shafts, etc., aiding mass producing, and Fredrick W. Taylor's contribution in the late 1800's, that of dividing each operation into elements and timed with a stop watch, enabled Management to predict the time required to perform an operation, contributing much toward increasing production. The 1900's witnessed Ford with his mass production techniques, and World War I helped to develop still further, greater productivity.

However, little thought was given to Methods Engineering until the late 1920's and more generally in the 1930's. Few explored this fertile field of reducing movements to basic elements, establishing Standard Data, developing cycles of operations with all or almost all unnecessary movements eliminated, reducing fatigue by re-positioning, utilizing most effectively the various body extremities for balance, co-ordination, greater productivity, and improving quality by Statistical Quality Control concepts.

Expanding the above, it has been found quite beneficial in various mass production industries where flow of material might either be homogeneous or heterogeneous to utilize the following procedure in initially setting out a new operation, whether it be an assembly or machining operation:

> After announcement is made that a new operation or machining process is under consideration, Methods Engineering, utilizing one of the newer Scientific approaches such as M.T.M. (Motion Time Analysis, A. B. Segur) or M.T.M. (Methods Time Measurement, Maynard, Stegemerten, and Schwab), steps in, and in conjunction with Engineering, trains the operator to perform the operation in the most efficient and prescribed manner, according to the Methods Engineer's or Methods Analyst's write-up. This of necessity, implies that fancy fixtures, jigs, solonoids, micro-switches, automatic trips, etc., are inadequate to do the entire job correctly. Data such as developed by M.T.A., M.T.M, or other similar widely accepted procedures must be utilized and all irrelevant motions, avoidable delays, pre-positioning, waits, etc., be reduced as much as possible, following the prescribed laws or procedures laid out by these techniques.
>
> When Methods Engineering advises that the operation is properly engineered, and the operator properly trained, Quality Control then moves in and, knowing

Management's desires and requirements regarding the quality needed from that operation to satisfactorily meet the consuming public's demands, or for subsequent operations, sets up the correct Statistical Quality Control chart procedures to evaluate quality output intensively, usually for the next two days. Experience has proven that 30 minute samplings, where $\bar{X}$-R, or $\bar{p}$ chart controls are utilized, for 32 successive checks, will provide adequate information as to whether the job should be certified for an incentive rate. Experience has shown conclusively, that an application of $\bar{X}$-R, or $\bar{p}$ chart, with control limits, is most forceful in providing the necessary picture needed, prior to setting a rate. If impractical to use either types of control charts mentioned above, Quality Control can set an A.Q.L. based on definite lot sizes to be processed, or definite quantities processed in a prescribed period. A minimum of 10 lots is required, and preferably 20, with a previously determined A.Q.L., statistically set U.C.L., L.C.L. and $\bar{p}$ calculated.

If a state of control exists, Quality Control with the knowledge that product quality is satisfactory for succeeding operations, certifies the operation to the Standards or Methods Engineering Department, as ready for an incentive rate, based on Standard Data, that precludes need for using a stop watch. Quality Control also notifies the Job Evaluations Department simultaneously, so that they can evaluate this operation knowing:

1. The quality being produced meets Management's needs.

2. The Method is the best, developed by an accepted Scientific technique utilizing Standard Data.

3. The operator has been scientifically trained by the above method, (and not while on a temporary piece-work rate set by a stop watch).

4. Equipment, machines or fixtures required to perform the job have been engineered properly.

Should Quality Control find the process average, or quality audit for the 32 check period not satisfactory, it immediately notifies Engineering, Methods, the Production Department Manager, and Planning (or Production Control), by a Refusal to Certify form listing defectives, percentages, and a detailed analysis, that is completely explanatory to those who must improve the operation further. Here, lack of training, poor material, or equipment,

may be at fault.

Usually within several days, Methods Engineering will request a re-check by Quality Control, advising it has made revisions that should reduce or eliminate defectives. If the re-check by Quality Control, indicates quality levels have been met, and maintained, Quality Control issues its form titled Quality Certification - Prior to Incentive Rate, giving a copy to all concerned, and sending attached to Method's Engineering's copy, all Quality Control charts telling the complete story.

The above procedure, naturally must have variations woven into it to be adaptable for the many different types of manufacturing establishments across the country. However, the principles have been proven to hold time after time, with widely diverse applications. Thus, it shies away from several weaknesses heretofore prevalent in industries that must utilize incentives, in that rates cannot be set arbitrarily by a stop watch; rates cannot be set by Time Study Engineering ignoring quality requirements for this and subsequent operations, and assuming the usual procedures of 100% screening later, (which actually, often are directly the result of the Time Study Engineer's rate).

The above system will automatically remove from the production foreman, set-up man, production supervisor, or maintenance mechanic, responsibility of engineering a job while it is on the production line. Too often in the past, when a new machine or operation is placed in the production line, supposedly ready to lend itself immediately to mass producing and mass assembling of components, it is found defective in numerous phases, and the poor production supervisor or foreman, needing production, pitches in and actually does the engineering and development to bring the job into control. Too often, a temporary or special rate is established. This rate, supposedly temporary, oft-times becomes somewhat more than just temporary and 100% screening is inevitable. Refusal to accept a defective job or machine of this nature by production, and charging any piece-meal labor from it to the Variation account on the Budget (but never to direct labor) usually has a salutary effect in bringing concrete, positive action.

We might point out at this time, two variations of the above relationship between Statistical Quality Control and use of Standard Data, as a basis for:

1. bonus payments to operators over and above their flat hourly rate,

2. as a basis for hourly-rated workers paid on the number of lots of material passing Quality Control acceptance

sampling.

Under No. 1, let's assume we have a department with many types of operations, few homogeneous, but each operation repetitive in itself, and handling many hundreds or perhaps a thousand or more items each day, not necessarily in specific lots, usually flowing progressively from one operation to another, all on a flat rate, and difficult to 100% screen.

Utilizing M.T.A., M.T.M. or Standard Data, with accompanying procedures for improving work methods, the hourly or daily quota is established, all motions in each job analyzed, reduced, combined, etc., as well as all necessary tools, fixtures, equipment pre-positioned, for streamlining the operation by prescribed Scientific techniques. Quality Control, then steps in, and treating the day's established quota as a grand lot, samples the operation hourly, using accepted, statistically correct sampling techniques and sizes, based on the hourly, or sub-lot sizes, determines the true process average, usually requiring 20 days for a most complete picture. This applies particularly to attributes sampling. Then, considering all conditions, and quality requirements for desired ultimate consumer acceptance, and in light of necessary costs to maintain normal profits, Quality Control establishes an A.Q.L., usually a minimum of 1% above the process average recently found. (If the process average is too high in Quality Control's opinion, the operation is completely rejected, and usually re-engineered).

Hourly, samplings are made at each operation, and the total percent found defective plotted daily. If workers are paid weekly, the process average is computed at the end of each week; if paid semi-monthly, the process average is computed every 15 days, but in either case, it is ruled plainly on the chart. If the process average is equal to or less than the A.Q.L. previously established, the operator is eligible for bonuses, based on the quantity with which he or she has exceeded the daily quota or work task.

For example, if a press assembler established a process average on his chart for the 5 day week as 1.2% and his A.Q.L. had been determined previously as 2%, he is eligible for bonuses. While his flat hourly rate might be $1.40 per hour, he might have exceeded his hourly rate by 40 pieces per hour, and according to the bonus scale previously prepared by Methods Engineering, in conjunction with Management, 40 pieces per hour is equal to (for example) 20¢ increment hourly. Thus, in an 8 hour day, the press assembler in this case would receive $1.60 per day for that week, over and above his flat hourly rate. Increments of increase would be worked out for all possible quantities. Other factors too must be considered but cannot be covered here, include:

A. The Rate Group.

B. Job Evaluation.

C. Failure to make base quota production.

D. Training Time.

E. Failure to continue at higher incentives, once established.

F. Action required when A.Q.L. exceeded.

G. Upper and Lower Control Limits.

Under No. 2 suggestion, it might be assumed that 100% screening is pratical and available. Also, we might assume if the quality level of complete lots off several similar operations, each performed by a different operator, is such as to warrant a Quality Acceptance Sampling Program, it by-passes 100% inspection. Included in the base rate for each operator is time allowed for re-working material rejected by 100% inspection. (This must be re-submitted for re-inspection after re-working).

Now - if material is accepted by the lot-by-lot Quality Control Sampler, it not only by-passes 100% inspection, but the operator can requisition more lots of material. Assuming he gets paid by the lots or pieces handled daily, he is in a position to process more material per day, and receive more pay, than one whose work is rejected by sampling, then sent to 100% inspection. This naturally slows down the operator, and prevents him from processing more lots. In other words, the higher his quality, the greater the probability of more lots passing Acceptance Sampling, the higher the quantity of material which he can process daily, for greater income.

Naturally, there are other details and preliminaries that must of necessity be worked out, but time will not permit detailing in this paper.

Suffice it to say, the above examples constitute a frame work of basic scientific Management principles for relating satisfactorily, Quantity and Quality, or Incentives and Quality Demands. Diligently persued, and honestly "engineered" by all concerned, <u>everyone</u> can benefit. Again, this writer feels, until we consummate this "marriage" between vigorous, practical, Statistical Quality Control, and the latest Scientific methods and techniques, we will not have achieved this ultimate, and most desirable goal.

ANALYSIS OF COVARIANCE

Irving W. Burr
Purdue University

Editor's note: Due to illness, Dr. Burr was unable to prepare his paper in time for inclusion here. Copies of his paper will be distributed at the Convention. Anyone else wanting a copy may write to Dr. Irving W. Burr, Statistical Laboratory, Purdue University, Lafayette, Indiana.

ESTABLISHMENT OF QUALITY LEVEL THRU COMPONENT AND SYSTEM TESTING

Ralph S. Reade
Consolidated Vultee Aircraft Corporation

I'm sure you have all heard the sage observation that "quality cannot be inspected into a product, it must be built there." Undoubtedly some of us have taken refuge behind this observation when called on the carpet to explain a rash of poor workmanship or an obvious lack of quality in the products your company produces.

The subject I am about to discuss, I believe, will show that quality goes much farther back than the making and inspecting phases.

An axiom which is often heard is the main theme of my paper. "An ounce of prevention is worth a pound of cure! How true this old axiom is and how frequently it can be applied to today's complex problems.

Before we get further into the subject it appears desirable to establish certain definitions so that there will be a common basis for understanding of some terms which will be used. These may or may not be peculiar to the aircraft industry.

1. Detail Part - A part which is the first step beyond the raw material stage. A part made exclusively from a type of material or combining several different materials to produce an item having a measurable function. Examples: wire, gaskets, insulation, electronic tube etc.
2. Sub Component - The assembly of a collection of detail parts into a semi-operable unit.
3. Component - A complex item completing the assembly of one or more sub-components into an operable unit.
4. Complex Item - An item made up of two or more detail parts. The more detail parts the greater the degree of complexity.
5. Environmental Tests - A series of specific tests designed to determine the degree of resistance of a detail part or complex item, to normal operating environments such as corrosion, humidity, fungus, vibration, sand and dust, altitude, operating temperatures, etc. With modern global operation of aircraft these tests as nearly as practical simulate the environments which aircraft will encounter on or above the earth. We are rapidly approaching the time when this will become impractical on extremely high speed, high altitude aircraft and missiles.
6. Complex System - A system made up of two or more complex items which must work together to accomplish a given function.
7. Integration Tests - Those tests required to prove compatability of complex items combined into a complex system. These tests are more critical where complex items from several manufacturers are combined into a complex system.

Before getting into the analysis of complex items and systems. A quick look at the steps of manufacture is undoubtedly in order.

First, materials are brought into being as a result of research. We will not dwell on this because I believe that it is a foregone conclusion that quality of applied research has a definite bearing on the quality of

of an end product. In materials research, however, frequently the researcher knows neither the quality nor the possible uses for the materials which are the result of his efforts. Many materials are brought into being as by products of a research program.

The second step in the manufacture of raw materials. This requires putting the results of research into useable form. At this point the first practical step in establishing the quality level of a companies end product begins. At this point either the raw material manufacturer or the purchaser should establish the characteristics and properties of the material if it is to be of maximum value to the engineer for design purposes.

The third step is the manufacture of detail parts from this raw material. Again this is the ideal time for the user to determine what material is best for the ultimate end use of the detail part.

Fourth the fabrication of detail parts into a sub-component. At this point the quality of the raw materials and detail parts begin to manifest themselves in the final design. To do quality design an engineer should be backed up by not only good materials and detail parts but also be reliable design information on these parts and materials.

The assembly of these sub-components into a complex item or component is usually the first time manufacturers like Convair get a look at them.

They are normally purchased to non-standard specifications which establish performance under all of the operating environments. We feel it is our responsibility to protect ourselves by conducting our own tests to the greatest possible extent. When it is necessary to allow the manufacturer to conduct the required tests, we demand the right to monitor his tests. Not that we do not trust our vendors implicity, we trust him to the same extent we do our most trusted bookkeeper whose books we do not allow him to audit himself. This testing, we feel, is paying extremely high dividends. It protects our customer, our vendor and ourselves.

Testing of a complex item begins when we receive engineering models and prototypes. These items are put through the series of tests required by the purchase specification. These tests are as realistic as our specification engineers know how to make them.

We usually start with 3 units and run the series of evaluation tests. A typical testing sequence is as follows:

Unit No. 1	Unit No. 2	Unit No. 3
Examination of Product	Examination of Product	Examination of Product
Case Leakage	Case Leakage	Case Leakage
Dielectric Test	Dielectric Test	Dielectric Test
Functional Test	Functional Test	Functional Test
Shock Test	Low Temperature Test	Endurance Test
	High Temperature Test	Aging Test
	Altitude Test	
	Thermal Shock Test	
	Salt Spray Test	

These tests prove pretty conclusively whether the design engineer had adequate knowledge of the detail parts or not. It also proves how much testing the manufacturer did himself before submitting his parts to these tests.

We'll take a few moments and discuss some of these evaluation tests.

Salt Spray - This test is supposed to determine the resistance to corrosion of this complex item. However, the standard salt spray test which is being universally used, as established by Government specification QQ-M-151 was developed primarily for the purpose of testing metal finishes It is of little use in determining the corrosion resistance of a component. Since it is a steady state test and there is no way for the salt atmosphere to penetrate to the inside of the part, there is no way to cause the corrosion except externally, which again is a function of the finish.

Humidity-Cycling - This test is of infinately greater value in determining corrosion resistance than the salt spray test. Cycling Humidity is run in moisture saturated atmosphere and the temperature cycled so that each cycle causes internal condensation. This test draws moist air through any pinhole or opening anywhere in the unit. Reducing the temperature causes internal condensation. This test detects three things.

1. How well the internal parts resist the presence of moisture.
2. Whether the moisture causes short circuiting of electrical contacts.
3. How familiar the designer was the effects of moisture on dissimilar metals combinations.

This humidity condition exists throughout the world in the temperate and tropic zones.

Fungus - This test determines how well the organic materials used in the design resist fungus growth or possibly a better way to express it would be, whether the organic materials used, support fungus growth. This test is made by spraying fungus spores on the parts containing insulations, fabrics, paper etc. and keeping them under warm, moist conditions for a specified time.

Sand and Dust - This is a test to determine the effects of air borne sand and dust particles on bearings and seals. Its obvious how this condition could effect the service life of an item which has external moving parts.

High and Low Temperature - These tests serve several purposes.

1. The effect of thermal change on the dissimilar metals. If improper materials were selected by the designer the unit may bind or sieze when taken from -65°F to the high temp requirement of 180°F or over. Both of these temperatures are, of course possible, at sea level, much higher temperature will be encountered in flight of supersonic aircraft and missiles.

2. The effect of both high and low temperature on lubricants. Finding lubricants which will withstand the ever widening range of temperatures is becoming increasingly difficult. From the lowest ground temperature of -65°F to the high operating temperatures of some of today's supersonic aircraft and missiles could be 300 to 400 degrees.

Altitude - Altitude sensivity is a problem in many types of equipment. Effective altitude test should combine rates of climb, temperatures and vibration resistance requirements.

You might say this is all very fine but how do you establish the quality level with these tests. Well these tests tell us many things.

1. How much effort the vendor has gone to, to determine the proper materials and detail parts.
2. The quality of his designers and their familiarity with the operational problems.
3. The effectiveness of his developmental testing program.
4. Reliability under expected operating conditions.

These tests do not assure a satisfactory production part, however, they do establish the capabilities of the manufacturer and his designers. How better can a level of quality be established than a series of destructive tests which can determine maximum performance in the expected operating environments? What better than a series of controlled scientific experiments to determine if production parts come up to the quality a manufacturer is capable of attaining. How better can a determination of quality deterioration be made than to occasionally pull out a production part and subject it to the same series of tests as these units which were supposed to be representative of a manufacturers capabilities?

To return to the axiom quoted earlier "An ounce of preventation is worth a pound of cure". What can happen if the ounce of prevention is not used? If a manufacturer of a shut-off valve buys a limit switch from a supposedly "qualified" parts manufacturer and assumes it is a satisfactory part without making any attempt to verify its performance for this new set of requirements, it will be built into the sub-component and into the component without being thoroughly investigated. By the time it gets to the ultimate user the manufacturer may have built and delivered 50 of these components before the tests are complete which may show the switch to be faulty for the purpose intended. If this part has been determined as satisfactory for a vibration frequency range of from 0 - 50 cycles per second and it is to be used in a jet aircraft where reliability from 0-500 cps is required it is fairly obvious what will happen. Remembering that this part was "qualified" and used in good faith by the designer, what can be the result of lack of preventive testing? The switch could probably have been tested through the required frequency range in an hours time. Detecting, determining the cause of malfunctioning, and correcting the cause in the sub-assembly stage could easily cost 50 man hours. If the malfunction is found by the buyer during his series of evaluation tests look what happens. 50 units may have been manufactured before the cause of trouble is found.

If the failure is sufficiently vital as to affect safety of the aircraft or missle, production could be halted until all parts have been redesigned, or at least replacement of the malfunctioning part has been achieved. If no survey of switches has been made to determine the best for the particular purpose, long delays in correcting the trouble may

ensure while the tests which should have been made before the design was started are accomplished. Especially in the aircraft industry where no lead time is allowed to accomplish the testing ahead of production, it is a serious problem which can only be solved by supreme effort on the part of all parties concerned. To do as much preventive work as possible, in the time allowed, and do it at the earliest possible date, should be the ultimate aim of all of us concerned with quality of aircraft.

Systems Tests - As previously defined, systems are made up of two or more complex items. It is highly impractical to integrate these items into a working system at any location other than the laboratories of the prime contractor. It is fairly obvious that parts purchased from many sources, which must be assembled with parts made in the prime contractors plant, can only be integrated into a working systems by the prime contractor.

In automobiles much of the testing for performance and durability is done at special proving grounds which have all of the hazards and environments of modern driving. In aircraft it is economically unsound to do this testing in the air under actual flight conditions. Consequently, as much systems integration testing as is possible should be conducted in laboratories equipped with altitude facilities, vibration, high and low temperatures and these facilities so arranged that any type of test desired or any combinations of conditions can be achieved.

You may ask why is it necessary to test these as a system if all of the individual parts have already been tested to the specification and found to be satisfactory? Lets take a look at this. The shut-off valve in the fuel system has never been tested with the actual fuel pump applying the pressure. It is highly conceivable that the characteristics of the pump could cause a malfunction of the valve. Neither has it been integrated with the various other valves in the system. Up to this point the compatability of the parts one with the other has not been established.

In order to determine how well this system will react to flight environments it must be tested in those environments. The system should be completely instrumented for pressure, flow, temperature, frequency of vibration, amplitude and any other factual data needed to make a complete evaluation. The system is subjected to the simulated flight conditions in an altitude chamber under flight altitudes, with the proper ambient temperatures and subjected to the proper range of frequencies. Of course, since the airplane has not yet flown, these conditions are assumed, using the latest available technical information.

If all of the tests come out satisfactorily we have established a quality level. Where is that quality level? We don't know, maybe we'll never know, whether it is barely passable or nearly perfection. If statistics are any proof we know that these aircraft parts are barely passable. How do I dare make such a statement? Because statistics show that of the components tested in the series of tests discussed under complex items, that 96% of those parts failed to meet the specification requirements during the first series of tests. They are, of course, modified and redesigned after failure until they will conform to specifi-

cation requirements.

What then is the answer? Preventive thinking and Preventive testing. The ounce of prevention! Is the effort made to prevent trouble in the proper proportion to the complexity and cost of a modern defensive weapon? Is the preventive effort out of proportion to the corrective effort? Many companies maintain "Corrective Action Committees". How many have "Preventive Action Committees"? Can foresight be used as effectively as hindsight?

COMPONENTS OF VARIANCE AND MIXED MODELS

R. L. Anderson
North Carolina State College

1. Introduction. In this paper, I wish to discuss that aspect of quantitative measurements which requires statistical treatment--variability. Individual products tend to vary, even when they come from the same production process. The identification of the sources of this variation and the development of methods of estimating the separate variances and testing hypotheses concerning them is the special province of the statistician. If a given product is being prepared for production, the statistician may be asked to cooperate with the production managers and research engineers to reduce the variability. In many cases, these estimates of variances are used to indicate how various sampling plans will affect the variability of sample averages. These sample averages may be used to place a quality index or grade on the product; they may be used in experiments to compare different production methods or different combinations of the factors of production.

There is a tendency to classify analyses of variation into three categories:

(i) All sources of variation except one are fixed in repeated sampling--the regression model or analysis of variance model.

(ii) All sources of variation are essentially random--the random model or variance component model.

(iii) There is a mixture of fixed and random sources of variation--the mixed model.

Actually this categorization is rather naive, since we have blends of random and fixed components in all research and production procedures. However I will confine myself to a discussion of the last two categories; the regression or analysis of variance model has been discussed by the previous speaker, and will be considered in other papers to follow.

I devoted four chapters of a book written in 1952 with T. A. Bancroft to the general theory of variance components. At that time I indicated a number of theoretical problems which needed to be tackled. I will mention several of these problems in this paper, because I feel that they are quite important and have been avoided because they do not present neat mathematical solutions. If anyone has statistical problems similar to those mentioned here, I would appreciate hearing of them. I do not promise solutions; however, if one can accumulate enough evidence of the need for solutions, perhaps the theoretical statisticians can be induced to think about them.

In attempting to delimit the subject matter of this paper, I have introduced some terms which may be unfamiliar to many. What is the distinction between random and fixed sources of variation? I will follow the usual practice of designating variables as fixed if the entire population about which references are to be made is in the sample or experiment. If the sample does not include the entire population, the variables are said to be random. Random variables are usually called variates. If the population is finite, but only a fraction of the population is covered in the sample, one may need to consider the usual finite

population correction factors* in his analysis; I will not introduce these correction factors in this paper. How do we know the sample is representative of the population? In many cases the variates used are selected because of convenience and not at random from the larger populations; this is especially true in many experiments with so-called random block or year effects. Since the only thing assumed to be random is the effect (and not the block or the year), one hopes that Nature has done a good job of randomly distributing her effects.

2. Random Model. In order to clarify these rather vague remarks, let me present some specific examples. Suppose we wish to estimate the sources of variation in sampling procedures to grade bales of wool. The sampling procedure is as follows: select n bales from the lot and then k cores from each bale. Each core or composite of several cores is subjected to laboratory analysis to determine the percent clean content, p. Suppose the value of p for the j-th core from the i-th bale can be represented as follows:

$$p_{ij} = P + b_i + c_{ij};\ i = 1,2, \ldots, n;\ j = 1,2, \ldots, k,$$

where P is the true percentage for the entire lot; b_i the deviation of the percentage for the i-th bale from P (the true percentage for the i-th bale is $P + b_i$); c_{ij} is the deviation of the percentage for the j-th core of the i-th bale from b_i. This model assumes that the deviations are actually additive. We also assume that the n bales are selected in some random manner from all bales (N) in the lot and the k cores at random from each sample bale. I doubt that anyone can rigorously define random; however, I believe that almost everyone understands what it means. One can number the bales in the lot from 1 to N; then he can use a table of random numbers or a box with slips numbered from 1 to N to select the n sample bales. Random selection of the k cores from each bale is more difficult, but various schemes are available.

If the core-to-core variability is the same for all bales, the analysis of variance technique can be used to estimate the variability between bales and between cores in the same bale. I shall use the variance as the measure of variability. The previous speaker has discussed the analysis of variance, which is as follows for the wool samples, if one neglects finite population correction factors.

Table 1. Analysis of Variance for Wool Data

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected value of Mean Square
Between bales	$n - 1$	SSB	MSB	$\sigma_c^2 + k\sigma_b^2$
Between cores in bales	$n(k - 1)$	SSC	MSC	σ_c^2

*The usual formula for the variance of the mean of n observations is σ^2/n; if n is selected from a population of size N, the variance of the mean is

$$\frac{N-n}{N} \frac{\sigma^2}{n} ,$$

which is approximately σ^2/n if N is large compared to n.

I am using SS to stand for sum of squares and MS for mean square. Now what is meant by the expected value of the mean square? This is the average value of the mean square over all possible samples of n bales and k cores per bale from the lot of N bales, in which the deviations are assumed to have variances σ_c^2 for cores and σ_b^2 for bales. If we computed the variance of the c_{ij} for each bale and then averaged over all bales in the lot, this average variance would be σ_c^2. Similarly the variance of the b_i for all N bales is σ_b^2. This analysis assumes all testing errors are contained in the c's. Methods are available to subdivide σ_c^2 into the two parts

$$(\sigma_c^2 = \sigma_s^2 + \sigma_t^2 \; ,$$

where σ_s^2 is sampling error and σ_t^2 is testing error).

Before proceeding to derive the expectations, I will use these summations:

$$B_i = \sum_j p_{ij} = k\,P + kb_i + \sum_j c_{ij}$$

$$G = \sum_i B_i = nk\,P + k\sum_i b_i + \sum_i \sum_j c_{ij}.$$

If the b's and c's are randomly selected, the average value of $(B_i - kP)^2$ is

$$k^2\sigma_b^2 + k\sigma_c^2$$

and the average value of $(G - nk\,P)^2$ is

$$nk^2\,\sigma_b^2 + nk\,\sigma_c^2.$$

It can be shown that

$$MSB = \frac{1}{n-1}\left\{\frac{\sum (B_i - kP)^2}{k} - \frac{(G - nk\,P)^2}{nk}\right\}.$$

Hence the average value of MSB over all possible samples is

$$\frac{(nk - k)\sigma_b^2 + (n-1)\sigma_c^2}{n-1} = k\sigma_b^2 + \sigma_c^2 \; .$$

I will call this V_b. Similarly

$$MSC = \frac{1}{n(k-1)}\left\{\sum_i \sum_j (p_{ij} - P)^2 - \sum_i \frac{(B_i - kP)^2}{k}\right\}.$$

Since the average value of $(p_{ij} - P)^2$ is $\sigma_b^2 + \sigma_c^2$, the average value of MSC is

$$\frac{(nk = nk)\sigma_b^2 + (nk - n)\sigma_c^2}{n(k-1)} = \sigma_c^2 \; .$$

In future discussions, V_c and σ_c^2 will be used interchangeably to mean the same thing; also s_c^2 and MSC.

Returning to the analysis of variance table, we see that the estimate of the core-to-core variance is the between core (in bales) mean square (MSB), and the estimate of the bale to bale variance is simply

$$s_b^2 = \frac{MSB-MSC}{k}.$$

These estimates can be used to set confidence limits on the true mean percentage, P, if the deviations are approximately normally distributed. Once sufficiently reliable estimates of σ_b^2 and σ_c^2 are available, the statistician can prescribe a future sampling plan to best estimate P. If the cost of pulling a bale out of the lot is c_1 and of core-boring and analyzing the core is c_2, the total cost for $\underline{n}$ bales and $\underline{k}$ cores per bale is

$$C = n\, c_1 + nk\, c_2.$$

The variance of a sample mean of $\underline{nk}$ cores is

$$\frac{\sigma_b^2}{n} + \frac{\sigma_c^2}{nk}.$$

The value of $\underline{k}$ to obtain the minimum variance of the mean for a fixed cost or the minimum cost for a fixed variance is

$$k = \frac{\sigma_c}{\sigma_b}\sqrt{\frac{c_1}{c_2}};$$

$\underline{n}$ is determined to satisfy the desired cost or variance. One chooses the nearest integral value of $\underline{k}$ which gives the lowest variance or cost. If k' is the integer smaller than $\underline{k}$, one uses $\underline{k}$' cores per bale, unless

$$\frac{\sigma_c^2}{\sigma_b^2} \cdot \frac{c_1}{c_2} > k'(k' + 1).$$

J. M. Cameron (1951) presents some data with n = 7, k = 4, MSB = 10.9938 and MSC = 6.2606. Hence

$$s_b^2 = \frac{10.9938 - 6.2606}{4} = 1.183,$$

and the ratio of sample standard deviations is $s_c/s_b = 2.3$. If $\sigma_c/\sigma_b = 2.3$ and $c_1/c_2 = 1/4$, the value of $\underline{k}$ should be

$$(2.3)/2 = 1.15.$$

In this case, k' = 1. Since

$$\frac{\sigma_c^2}{\sigma_b^2} \cdot \frac{c_1}{c_2} = \frac{5.3}{4} = 1.3 < 2,$$

one should use one core per bale.

The use of variance components to plan future sampling reveals a need to estimate these components as precisely as possible. Cameron discusses methods of setting up experiments especially designed to estimate the variance components. Since it is quite expensive to analyze each core separately, it is customary to composite all k cores from a bale and run one analysis on the composite sample. Unfortunately this procedure does not provide the necessary data to estimate σ_b^2 and σ_c^2. One procedure is as follows: assume we have 2 M lots with n bales per lot and k_1 cores per bale for M lots and k_2 cores per bale for the other M lots. All cores for each lot are composited. If we let V_1 be the mean square between the M lots with nk_1 cores per lot and V_2 for the other M lots, σ_c^2 is estimated by

$$\frac{nk_1 k_2}{k_2 - k_1} (V_1 - V_2).$$

If σ_a^2 is the lot-to-lot variance component, $(n\sigma_a^2 + \sigma_b^2)$ is estimated by

$$\frac{n}{k_2 - k_1} (k_2 V_2 - k_1 V_1).$$

Another procedure suggested by Tanner enables one to obtain an estimate of σ_b^2.

If the deviations (b_i and c_{ij}) are drawn from normal populations, the variances of the estimated variance components are simple functions of the true variance components. If any estimated component is a linear function of mean squares in the analysis of variance, i.e.

$$s^2 = a_1 MS_1 + a_2 MS_2 + \ldots + a_r MS_r ,$$

the variance of s^2 is simply

$$\text{Var}(s^2) = 2 \sum \frac{a_i^2 V_i^2}{(df)_i} ,$$

where V_i is the expected value of MS_i and $(df)_i$ is the number of degrees of freedom for MS_i. For example in Table 1, the variance of s_c^2 (the estimate of σ_c^2) is simply $2V_c^2/n(k-1)$ and the variance of s_b^2 (the estimate of σ_b^2 is

$$\text{Var}(s_b^2) = 2\left[\frac{V_b^2}{n-1} + \frac{V_c^2}{n(k-1)}\right] / k^2.$$

The variance formulas are only slightly more complicated when compositing is done. Hence if one knows the cost ratio (c_2/c_1), he can evaluate

various sampling plans as to their usefulness in estimating the variance components. Unfortunately these variance formulas may be in serious error if the parent distributions are not normal; some research is being carried on in this field, but I feel that we have barely scratched the surface.

In order to illustrate other problems in the estimation of variance components, I would like to discuss a problem which was present to me at Purdue University in 1950. An experiment was to be designed to test various strains of bacteria for their efficiency in the production of streptomycin. Before setting up this experiment, a preliminary experiment was to be conducted to estimate the magnitudes of the various sources of variability in the production and assay process. There were five stages in this process: an initial incubation stage in a test tube, a primary inoculation period, a secondary inoculation period, a fermentation period in a bath and the final assay of the amount of streptomycin produced. The variance components for the variation at each stage of sampling were (starting with the incubation):

$$\sigma_t^2,\ \sigma_p^2,\ \sigma_s^2,\ \sigma_f^2 \text{ and } \sigma_a^2 .$$

Previous experience indicated that σ_t^2 and σ_p^2 were very important sources of variation; hence, the sampling plan should provide for good estimates of these components. Suppose 80 assays could be made. The most obvious sampling plan would be to use 5 test tubes, draw 2 samples from each test tube for the primary inoculation, 2 samples from each primary for the secondary inoculation, 2 samples from each secondary for the fermentation, and assay 2 samples from each fermentation bath. Unfortunately this balanced sampling procedure provides only 4 degrees of freedom to determine MST, which is the only mean square containing σ_t^2 in its expectation; similarly there are only 5 degrees of freedom to determine MSP. But there are 40 degrees of freedom to estimate σ_a^2, which is usually quite small.

In order to increase the degrees of freedom for MST and MSP, one could consider a staggered sampling plan, such as the following: Use 2 test tubes as above; 2 test tubes as above, except only 1 assay per fermentation; 4 test tubes as above, except only one fermentation per secondary and 1 assay per fermentation; 8 test tubes with two primaries each and 1 of each stage thereafter. The analysis of variance has 15 degrees of freedom for MST and 16 degrees of freedom for all other mean squares. The expectations of the mean squares, E(MS), are decidedly different in the two cases as is seen in Table 2.

Table 2. Expectations of Mean Squares for Streptomycin Experiment

	Balanced Design						Staggered Design					
		Coeff. of σ_i^2 in E(MS)						Coeff. of σ_i^2 in E (MS)				
Source of Variation	D.F.	σ_a^2	σ_f^2	σ_s^2	σ_p^2	σ_t^2	D.F.	σ_a^2	σ_f^2	σ_s^2	σ_p^2	σ_t^2
Test Tubes	4	1	2	4	8	16	15	1	1.107	1.44	2.36	4.72
Primary	5	1	2	4	8		16	1	1.125	1.50	2.50	
Secondary	10	1	2	4			16	1	1.250	2.00		
Fermentation	20	1	2				16	1	1.500			
Assay	40	1					16	1				
Total	79						79					

With the balanced design, each variance component is estimated by subtracting from its own mean square the one below it, e.g.

$$s_p^2 = (MSP-MSS)/8.$$

Hence the variance of the estimate is simply

$$\text{Var}\ (s_p^2) = 2\left[\frac{V_p^2}{5} + \frac{V_s^2}{10}\right] / 64 = \frac{V_p^2}{160} + \frac{V_s^2}{320},$$

where $V_p = E(MSP) = \sigma_a^2 + 2\sigma_f^2 + 4\sigma_s^2 + 8\sigma_p^2$ and $V_s = E(MSS) = \sigma_a^2 + 2\sigma_f^2 + 4\sigma_s^2$.

The estimation problem is much more complicated with the staggered design, e.g.

$$s_p^{*2} = (8MSP^* - 6MSS^* - MSF^* - MSA^*)/20$$

$$\text{Var}\ (s_p^{*2}) = 2\left[\frac{64V_p^{*2}}{16} + \frac{36V_s^{*2}}{16} + \frac{V_f^{*2}}{16} + \frac{V_a^{*2}}{16}\right] /400$$

$$= \frac{V_p^{*2}}{50} + \frac{V_s^{*2}}{88.9} + \frac{V_f^{*2} + V_a^{*2}}{3200},$$

where the starred (*) values refer to the analysis of the staggered design. A comparison of these two designs is difficult, because the expected values of the mean squares (V and V*) are not equal. If all the variance components were equal, Var (s_p^{*2}) would be only about 60% as large as Var (s_p^2). If σ_p^2 is larger than the other components, the staggered design is even more superior; however, if σ_p^2 is much smaller than the other components, the balanced design may be better. Of course there are an innumerable number of possible designs to be considered. This is another field that has scarcely been touched by research workers. Since the determination of the best plan depends so heavily on a knowledge of the relative magnitudes of the components, perhaps the estimation should be done sequentially. That is, one would conduct a preliminary experiment to obtain rough estimates of the relative magnitudes of the components, and then design a more elaborate experiment based on these results. All of this discussion points up the need for accumulating information on variance components, as research and production are in progress.

3. Mixed Model. Now for some examples of Mixed Models; first, an experiment conducted by W. C. Hackler of the University of Florida to compare the absorption properties of 15 ceramic compositions. In order to determine if these compositions responded in a similar manner regardless of the temperature under which they were fired, each composition was fired under three temperatures. Two batches of each composition were prepared, giving a total of 30 batches. Two firings were made at each of the three temperatures, giving a total of 6 firings. A sample was selected from each of the 30 batches for each firing, i.e. all batches were represented on each firing. Hence there was a total of 180 absorption measurements in the experiment, obtained as follows:

(i) Each fired specimen is boiled in water, during which it absorbs water.

(ii) It is taken from the water and weighed at regular intervals. When successive weights level off, the boiling is terminated.

(iii) The percentage increase over the weight after firing is the recorded absorption measurement.

The 180 measurements are given in Table 3.

Table 3. Percent Absorptions for 15 Ceramic Compositions Using 3 Different Firing Temperatures.

		Temperature 1		Temperature 2		Temperature 3		
		Firing		Firing		Firing		
Composition	Batch	1	2	1	2	1	2	Total
1	1	12.57	12.59	7.14	7.31	3.84	3.43	46.88
	2	13.63	13.39	7.87	7.90	4.35	5.15	52.29
2	1	14.25	14.23	8.21	8.31	4.37	4.79	54.16
	2	15.44	15.46	9.22	9.51	5.66	6.60	61.89
3	1	13.88	13.93	7.14	7.18	3.35	4.54	50.02
	2	13.21	13.16	7.37	7.33	3.52	4.29	48.88
4	1	13.59	13.71	7.63	7.51	3.89	3.91	50.24
	2	13.49	13.78	7.60	7.12	3.64	3.58	49.21
5	1	13.17	12.33	6.72	7.02	4.15	3.56	46.95
	2	12.94	12.99	6.92	6.70	3.80	3.70	47.05
6	1	11.35	11.73	6.20	6.41	3.82	3.79	43.30
	2	11.71	12.04	6.43	6.66	3.94	3.78	44.56
7	1	13.55	13.25	7.66	7.69	4.40	5.51	52.06
	2	13.26	13.19	7.81	7.97	6.03	6.08	54.34
8	1	11.48	11.41	5.98	6.21	3.52	3.59	42.19
	2	11.07	11.37	5.70	6.00	3.52	3.67	41.33
9	1	11.30	11.23	6.20	6.18	3.59	3.93	42.43
	2	11.51	11.39	5.97	5.82	3.65	3.31	41.65
10	1	15.65	15.87	8.85	8.83	4.12	5.17	58.49
	2	15.83	15.73	8.61	8.99	4.74	4.18	58.08
11	1	10.55	10.87	5.15	5.61	3.21	3.56	38.95
	2	10.81	10.57	5.27	5.34	3.24	3.69	38.92
12	1	12.90	13.13	7.91	7.89	4.69	4.59	51.11
	2	13.02	13.42	7.96	8.06	5.40	4.84	52.70
13	1	12.24	12.29	6.20	6.74	4.39	3.55	45.41
	2	13.33	13.28	7.49	8.01	4.00	4.49	50.60
14	1	12.39	12.72	6.76	7.06	4.27	3.88	47.08
	2	13.09	13.04	7.09	7.34	4.16	4.39	49.11
15	1	12.37	12.35	6.74	6.84	3.89	3.34	45.53
	2	12.27	12.55	6.74	6.97	3.83	3.30	45.66
Total		385.85	387.00	212.54	216.51	122.98	126.19	1451.07

Let p_{gihj} be the percent absorption for the j-th batch of the h-th composition on the i-th firing at the g-th temperature; i, j = 1,2; h = 1,2, ..., 15; g = 1,2,3. The mathematical model for this percentage is

$$p_{gihj} = P + T_g + C_h + (TC)_{gh} + f_{gi} + b_{hj} + (Tb)_{ghj} + (Cf)_{hgi} + e_{gihj}$$

P represents the true average percentage absorption for these 15 compositions fired at these 3 temperatures. T and C are fixed constants which represent the added effect (above the mean, P) of the particular temperature and composition, and (TC) measures the fixed interaction of T and C.

The remaining terms represent random effects: f, the variation between the results on two firings at the same nominal temperature; b, the variation between the results of the two batches of the same composition; (Tb) and (Cf), interaction effects between a fixed and a random variable; e, the remaining measurement and experimental variability. These random effects will be assumed to have respective variances of

$$\sigma_f^2,\ \sigma_b^2,\ \sigma_{Tb}^2,\ \sigma_{Cf}^2 \text{ and } \sigma_e^2.$$

As I indicate in my book, (Tb) and (Cf) are random in one direction only. Hence σ_{Tb}^2 and σ_{Cf}^2 do not appear in the expectations of the (Batch) and (Firings) mean squares in the Analysis Variance for these data in Table 4.

Table 4. Analysis of Variance for Data in Table 3.

Source of Variation	D.F.	Mean Square	Expected Value of Mean Square
Temperatures (T)	2	1179.99**	$\sigma_e^2 + 30\sigma_f^2 + 2\sigma_{Tb}^2 + 30\,(\sum T^2)$
Compositions (C)	14	10.34**	$\sigma_e^2 + 6\sigma_b^2 + 2\sigma_{Cf}^2 + 12(\sum C^2)/14$
T x C	28	1.113**	$\sigma_e^2 + 2\sigma_{Tb}^2 + 2\sigma_{Cf}^2 + \sum\sum (TC)^2/7$
Firings (in Temp.)	3	.1521	$\sigma_e^2 + 30\sigma_f^2$
Batches (in Comp.)	15	.7405**	$\sigma_e^2 + 6\sigma_b^2$
T x batches	30	.0857	$\sigma_e^2 + 2\sigma_{Tb}^2$
C x firings	42	.0818	$\sigma_e^2 + 2\sigma_{Cf}^2$
Residual	45	.0631	σ_e^2

**Significant at 1% probability level.

In this particular experiment, it was expected that neither σ_{Tb}^2 nor σ_{Cf}^2 would be very large because the firings and batches were handled in a uniform manner. σ_{Tb}^2 would be large only if the two batches of the s same composition were enough different so that they would react differently to a given temperature. Similarly σ_{Cf}^2 could be large only if the firings at nominally the same temperature tended to have decidedly differ- temperatures, and there was a real (TC) interaction; in general, the two firings for a given temperature were made at nearly this temperature.

The analysis of variance in Table 4 assumes the typical form with the random components represented by variances, e.g. σ_e^2, and the fixed components by sums of squares of the constants. Estimates of the variance components are obtained as shown for the wool data, e.g. σ_{Cf}^2 is

estimated by

$$s^2_{Cf} = (.0818 - .0631)/2 = .0094.$$

Estimates of the P, T, C and (TC) constants are obtained by the usual regression procedure, as presented in any theory of the analysis of variance, e.g.

$$P = \frac{1451.07}{180} = 8.062; \quad T_1 = \frac{772.85}{60} - P = 4.819;$$

$$C_1 = \frac{99.17}{12} - P = 0.202; \; (TC)_{11} = \frac{52.18}{4} - (C_1 + T_1 + P) = -0.038$$

An examination of the last three mean squares in Table 4 confirms the expectation that σ^2_{Tb} and σ^2_{Cf} would be quite small. One can test the hypothesis that each is zero by the F-test. For σ^2_{Tb}, $F = .0857/.0631 = 1.36$, with 30 and 45 degrees of freedom; for σ^2_{Cf}, $F = .0818/.0631 = 1.30$, with 42 and 45 degrees of freedom. Neither F-value is significant at even the 10% level of significance. We note that there is an important batch to batch component of variation (σ^2_b), but that σ^2_f is not very important. Hence in future experiments it might be desirable to use more batches of each composition but only one firing at each temperature.

Since σ^2_{Tb} and σ^2_{Cf} are of trivial importance, I have neglected them in testing the fixed effects. It is obvious that all of these effects are significant at the 1% level. The temperatures used were purposely far apart; hence, their effects were almost certain to be different. With such large temperature effects, it is not surprising that the composition effects were somewhat different for the 3 temperatures, as shown by the significant (TC) interaction. You might ask how to make these tests if σ^2_{Tb} and σ^2_{Cf} could not be neglected. In this case there would be no single mean square to use as an error term. I discuss various methods of handling this problem in my book. I would use as an error term for T x C,

$$.0857 + .0818 - .0631 = .1044,$$

with the degrees of freedom approximated by

$$\frac{(.1044)^2}{\dfrac{(.0857)^2}{30} + \dfrac{(.0818)^2}{42} + \dfrac{(.0631)^2}{45}} = 22.$$

Hence one would obtain as an approximate F-statistic,

$$F \simeq 1.113/.1044 = 10.66,$$

with 28 and 22 degrees of freedom. The procedure is to add and subtract mean squares until the expected value of the result is the same as the error part of the mean square being tested. W. G. Cochran (1951) proposes a somewhat better procedure in which you add .0631 to the numerator of F, instead of subtracting from the denominator. Hence,

$$F \simeq 1.176/.1675 = 7.02,$$

with the following degrees of freedom:

$$\frac{(1.176)^2}{\dfrac{(1.113)^2}{28} + \dfrac{(.0631)^2}{45}} = 31 \text{ and } \frac{(.1675)^2}{\dfrac{(.0857)^2}{30} + \dfrac{(.0818)^2}{42}} = 69.$$

The Cochran procedure involves more computing and seems to change the results very little.

I want to conclude this paper with an example of a set of possible data which seem to be typical of some I have seen. Suppose that $\underline{p}$ shipments of wool are being compared, with cores being collected and tested by $\underline{r}$ different technicians. One would like to know if the shipments are essentially alike and if the technicians are obtaining essentially the same results for a given bale of wool. Inferences are to be made about only these $\underline{p}$ shipments and these $\underline{r}$ technicians. Assume that $\underline{n}$ bales are selected from each shipment, (giving a total of $\underline{pn}$ bales) and each technician selects and analyzes $\underline{k}$ different cores from each bale (giving a total of $\underline{rk}$ cores per bale).

Let p_{ghij} represent the estimated percentage of clean wool when the g-th technician analyzes his j-th core from the i-th bale of the h-th shipment. The mathematical model is

$$p_{ghij} = P + T_g + S_h + (TS)_{gh} + b_{hi} + (Tb)_{ghi} + c_{ghij}.$$

P, T, S, and (TS) are fixed components; b, (Tb) and c are random components with respective variances, σ_b^2, σ_{Tb}^2 and σ_c^2. Since personal judgement may have some influence in the selection and analysis of the cores, σ_{Tb}^2 may be an important source of variation in this experiment, as contrasted to the small values of σ_{Tb}^2 and σ_{Cf}^2 for the ceramics example. The Analysis of Variance is outlined in Table 5.

Table 5. Analysis of Variance for Extended Wool Experiment

Source of Varition	D.F.	Expectation of Mean Square
Technicians (T)	r-1	$\sigma_c^2 + k\sigma_{Tb}^2 + pnk\left[\sum T^2/(r-1)\right]$
Shipments (S)	p-1	$\sigma_c^2 + rk\sigma_b^2 + rnk\left[\sum S^2/(p-1)\right]$
T x S	(r-1)(p-1)	$\sigma_c^2 + k\sigma_{Tb}^2 + nk\left[\sum (ts)^2/(r-1)(p-1)\right]$
Bales (in Shipments)	(n-1) p	$\sigma_c^2 + rk\sigma_b^2$
T x bales	(r-1)(n-1)p	$\sigma_c^2 + k\sigma_{Tb}^2$
Cores (in bales)	(k-1) prn	σ_c^2

In many cases it is desirable to subdivide the sum of squares for cores in bales into a separate sum of squares for each technician with (k-1)pn degrees of freedom each. A comparison of the associated mean squares will indicate if the technicians have about the same variability in selecting and testing the cores. In this case, it would be useful to superimpose a compositing type of design in order to separate the sampling and testing components of σ_c^2.

In Table 5, the $\underline{r}$ technicians might be regarded as a sample of all technicians, i.e. the T components are random. In this case $\sum T^2/(r-1)$ is replaced by σ_T^2, and

$$\sum (TS)^2/(r-1)(p-1)$$

by σ_{TS}^2; also, σ_{TS}^2 appears in the expectation of the mean square for shipments. In addition, $k\sigma_{Tb}^2$ appears in all the expectations of all the mean squares except that for cores.

Finally if the shipments are regarded as a sample from a large number of shipments, we have a completely random model.

$$\sum s^2/(p-1)$$

is replaced by σ_S^2, and σ_{TS}^2 appears in the expectation of the mean square for technicians. The analysis of variance for this random model with interactions is given in Table 6.

Table 6. Analysis of Variance for Extended Wool Experiment: All Effects Random.

Source of Variation	Expectation of Mean Square
Technicians (T)	$\sigma_c^2 + k\sigma_{Tb}^2 + nk\sigma_{TS}^2 + pnk\ \sigma_T^2$
Shipments (S)	$\sigma_c^2 + k\sigma_{Tb}^2 + rk\ \sigma_b^2 + nk\sigma_{TS}^2 + rnk\sigma_S^2$
T x S	$\sigma_c^2 + k\sigma_{Tb}^2 + nk\sigma_{TS}^2$
Bales (in Shipments)	$\sigma_c^2 + k\sigma_{Tb}^2 + rk\sigma_b^2$
T x bales	$\sigma_c^2 + k\sigma_{Tb}^2$
Cores (in bales)	σ_c^2

Another version of this type of sampling was a problem sent to R. J. Hader of our staff last year as follows: Four separate bulk samples are taken from each of two batches of nominally similar material. These are randomly assigned, within each batch, to four different laboratories. Upon their arrival at the laboratories, these samples are each divided in some manner into two parts and these parts assigned randomly to two analysts. The two analysts should be randomly chosen from all the analysts at the given laboratory. The analysts then run duplicate determinations on their samples. The expectations of the mean squares are the same as in Table 6, with Cores replaced by Duplicates, Technicians by Batches, Shipments by Laboratories and Bales by Analysts; $k = n = r = 2$ and $p = 4$.

In many cases, inferences will be made about only these laboratories; in other cases, about only the given two analysts at each laboratory. The main point to remember is that the expectations of the mean squares, which in turn determine how to estimate the components, depend on the type of inferences one wants to make and on the sampling procedure.

I have omitted reference to many even more complicated problems in variance component analysis, especially the complications of a mixed model when the sampling is not balanced, such as with the staggered design for the streptomycin experiment. For example, consider the problem of estimating the differences between several strains of bacteria and between a number of laboratories, with the experiments for each strain at each laboratory following a staggered design. This and similar problems have many unsolved aspects regarding best estimating procedures.

Reference Cited

Anderson, R. L. and T. A. Bancroft (1952), Statistical Theory in Research, McGraw Hill Book Company, New York.

Cameron, J. M. (1951), "Use of Components of Variance in Preparing Schedules for the Sampling of Baled Wool," Biometrics, 7:83-96.

Cochran, W. G. (1951), "Testing a Linear Relation Among Variances," Biometrics, 7:17-32.

Newton, R. G., M. W. Philpott, H. F. Smith and W. G. Wren (1951), "Variability of Malayan Rubber," Ind. Eng. Chem., 43:329-334.

RESEARCH, DEVELOPMENT, PRODUCTION AND THE INSPECTION OF PRODUCT

Cdr. F.W. Russe, Jr.
Bureau of Ordnance, U. S. Navy

The Bureau follows the system laid out here in planning for the inspection of ordnance material purchased from industrial concerns. The plan also includes surveillance of material in the supply system owned by the Government.

This is an outline only, which will be expanded and detailed when presented.

The inspection planning system is based on the division of evolving equipment into several stages. These are distinct, yet the plan is flexible to permit changes as conditions change.

The stages of evolution are: A. research, B. development, C. evaluation, D.service, and E. withdrawn. Each is separated from the next by a specific decision.

It is necessary to consider three factors as basic. The end use of the item to be inspected must be designated. The item must be classed as either expendable or reuseable. The stage of evolution must be determined and specifically related to the material to be inspected.

For each stage of evolution the requirements of inspection are set forth. For A and B stages the only inspection to be done is that required by the design agency. Tentative Classifications of Defects are required to be used for material inspection during stage C. Production of material for service, which is stage D, calls for the use of formalized OCD. These are obtained for the inspector one month prior to the formal "release to production" of the design. Certain exceptions can be made, especially for reuseable material.

The OCD's are based on the specifications and each defect is listed only when it contributes to performance, interchangeability, safety, life, or coordination of the equipment or component being inspected.

Material in storage is continuously evaluated to determine its condition and availability for service use. The Ordnance Quality Evaluation Laboratories are used for this. The criteria include OCD's, specifications and information on deterioration.

The overall objective is to obtain material for service use by the Navy which is good and reliable and to do this most economically and logically.

Scheduling the evolution of equipment in this way permits sound planning for the training of personnel, preparation for activity in certain areas, and the balancing of inspection pressures.

Emphasis must be placed on the need for flexibility and the active re-scheduling and re-staging as conditions change.